

UNIVERSIDAD AUTÓNOMA DE MADRID

FACULTAD DE PSICOLOGÍA

Departamento de Psicología Social y Metodología



TESIS DOCTORAL

Contribuciones a la Teoría de Respuesta al Ítem no
Paramétrica.

Mario Luzardo

Directores:

Dr. Julio Olea

Dr. Vicente Ponsoda

Madrid, 2017

In memoriam, a mi madre, Elsa Verde.

Agradecimientos

Quiero agradecer en forma especial a mis directores de tesis Dr. Julio Olea y Dr. Vicente Ponsoda por su apoyo, paciencia y dirección, no solo durante el trabajo de tesis sino en todo mi proceso de estudios en la Universidad Autónoma de Madrid.

Destaco del Dr. Julio Olea, además de su excelente apoyo académico, su ayuda en todos los trámites administrativos haciéndose cargo de asuntos más allá de lo esperado. Estoy en deuda con él y espero algún día poder retribuirle.

Agradezco también al programa de científicos visitantes de la Universidad de la República de Uruguay, que trajo a la Facultad de Psicología de mi país al Dr. Vicente Ponsoda, permitiéndome tener contacto con este programa de doctorado.

Más importante aún fue que, a partir de esa visita, se comenzó a discutir con mayor énfasis la importancia de la enseñanza de la metodología y la psicometría en nuestra facultad.

Los cursos recibidos durante el doctorado han propiciado la transferencia de conocimiento desde la Facultad de Psicología de la Udelar a otras instituciones. Prueba de ello es la participación en el desarrollo del Sistema de Evaluación de Aprendizajes (SEA) de la Administración Nacional de Educación Pública, en la aplicación de la Prueba Adaptativa de Inglés desarrollada por el Plan Ceibal (programa OLPC) y los convenios de asesoría con el Instituto Nacional de Evaluación Educativa y con la Fundación Ceibal. Desde ya nuestro agradecimiento a la Universidad Autónoma de Madrid porque este programa de doctorado ha tenido gran influencia en posibilitar esto.

Finalmente, agradezco muy especialmente a mis padres y a mi esposa por darme lo más importante en mi vida: su incondicional cariño.

Índice general

Contenido	i
Introducción	1
1. Teoría de respuesta al ítem	6
1.1. Orígenes y primeros desarrollos	6
1.2. Supuestos de la teoría de respuesta al ítem	10
1.3. Modelos para ítems dicotómicos	13
1.3.1. Modelo de Rasch	15
1.3.2. Función de información	16
1.4. Modelos para ítems politómicos	17
1.4.1. Modelo de categorías nominales	18
1.4.2. Modelo de escala de clasificación	20
1.4.3. Modelo de respuesta graduada	21
1.4.4. Modelo de crédito parcial y modelo de crédito parcial generalizado	22
1.5. Modelos de respuesta continua	24
1.6. Estimación de modelos dicotómicos	25
1.7. Modelos multidimensionales	32
1.8. Otros modelos	35
1.9. Modelos de desdoblamiento	37
2. Modelos no paramétricos de teoría de respuesta al ítem	39
2.1. Introducción	39
2.2. Modelos de Mokken	40
2.2.1. Modelo monótono homogéneo	40
2.2.2. Modelo de doble monotonicidad	41
2.2.3. Estimación de la habilidad	42
2.3. Estimación bayesiana no paramétrica	44
2.3.1. Estimación usando deconvolución	45
2.4. Modelo no paramétrico unidimensional usando regresión no paramétrica	47
2.5. Enfoque desde el análisis de datos funcionales	50
2.5.1. Estimación invariante de la habilidad	50
2.5.2. Métricas	51

2.5.3. Algoritmo para el análisis funcional de un test	51
3. Modelo no paramétrico multidimensional	53
3.1. Introducción	53
3.2. Presentación del modelo multidimensional	54
3.2.1. Estimación de la CCI	56
3.2.2. Respuestas politómicas y continuas	58
3.3. Estimación de la densidad conjunta del rasgo	58
3.3.1. Caso del rasgo multidimensional	62
3.4. Núcleos	67
3.5. Elección del ancho de banda para densidad	70
3.6. Consistencia conjunta del modelo multivariado	71
3.6.1. Preliminares	72
3.6.2. Sucesión de funciones asintóticamente Lipschitz	73
3.7. Hipótesis para la consistencia conjunta	74
3.8. Prueba de la consistencia conjunta	75
3.9. Propiedades del estimador de la CCI si la habilidad es medida sin error	85
3.9.1. Distribución asintótica de la CCI	90
3.10. Sesgo y varianza del estimador multidimensional	92
3.10.1. Métodos de selección de ventanas	93
4. Modelo isótono	94
4.1. Modelo isótono	96
4.2. Propiedades del estimador cuando el rasgo es medido sin error	97
5. Simulaciones	110
5.1. Método	111
5.1.1. Medidas de ajuste de los rasgos	112
5.1.2. Distancia entre curvas características del ítem	112
5.1.3. Modelos de TRI utilizados	112
5.1.4. Generación de datos	113
5.2. Estudio 1: Ajuste en la estimación del rasgo.	115
5.3. Estudio 2: Distribución de la estimación del rasgo	119
5.4. Estudio 3: Efecto del tamaño muestral en la estimación de las CCI unidimensionales.	121
5.5. Estudio 4: Tamaño muestral en dos dimensiones. Caso I.	122
5.6. Estudio 5: Tamaño muestral en dos dimensiones. Caso II.	123
5.7. Estudio 6: Ajuste por ítems y sujetos en dos dimensiones.	124
5.8. Estudio 7: Efecto de la elección de la ventana en una dimensión.	125
5.9. Estudio 8: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso I.	127
5.10. Estudio 9: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso II.	128

5.11. Estudio 10: Efecto del tipo de núcleo en una dimensión.	128
5.12. Estudio 11: Efecto de distintos núcleos en dos dimensiones. Caso I.	130
5.13. Estudio 12: Efecto de distintos núcleos en dos dimensiones. Caso II.	130
5.14. Estudio 13: Efectos de la distribución de los rasgos en una dimensión	131
5.15. Estudio 14: Efecto de distintas distribuciones marginales de θ en dos dimensiones.	
Caso I.	132
5.15.1. Distintos niveles de correlación entre los rasgos	132
5.16. Estudio 15: Efecto de distintas distribuciones de θ en dos dimensiones. Caso II. .	133
5.16.1. Efecto de distintos grados de correlación entre los rasgos	133
5.17. Estudio 16: Efecto de la elección de la ventana en el modelo isótono	133
5.18. Estudio 17. Efecto de la elección del núcleo en el modelo isótono	135
6. Aplicación a datos reales	137
6.1. Objetivos	137
6.2. Métodos	137
6.3. Evaluación diagnóstica en matemática	138
6.3.1. Evidencias sobre la estructura interna	139
6.4. Estimación no paramétrica en una dimensión	143
6.5. Estimación no paramétrica en dos dimensiones	144
6.6. Aplicación al Children Depression Inventory	145
6.7. Estimación multidimensional	147
7. Conclusiones	149
7.1. Aportes de la tesis	149
7.2. Limitaciones y problemas abiertos	153
Referencias	154
8. Anexo I Gráficas de las Aplicaciones	168
8.1. CCI del cuadernillo 1 de matemática	168
8.2. Superficies estimadas en dos dimensiones	171
8.3. FRI de las dimensiones del CDI	174
8.4. Proyecciones bidimensionales	176

Índice de figuras

3.1. Proximidad de la densidad real y estimada por tipo de distribución	59
3.2. Principales núcleos unidimensionales	68
3.3. Núcleos multiplicativos de dimensión 2	69
5.1. Ajuste por modelo	116
5.2. Ajuste por cantidad de sujetos	116
5.3. Ajuste por cantidad de sujetos	117
5.4. Densidad y cópula del rasgo	119
5.5. DT por ítems y sujetos	120
5.6. DT y estimado	120
5.7. Intervalos de confianza	121
5.8. Ajuste y tamaño muestral	122
5.9. Modelo real	123
5.10. Estimado	123
5.11. Real	124
5.12. Estimado	124
5.13. CCI para diferentes ventanas	127
6.1. Muestreo de la prueba	139
6.2. Prueba 1-Cantidad de factores	140
6.3. Prueba 2- Cantidad de factores	141
6.4. CCI cuadernillo 1	141
6.5. Items 1,5,9 y 13	141
6.6. CCI cuadernillo 2	141
6.7. Estimaciones de la CCI ítem 13	144
6.8. Densidad de la habilidad	145
6.9. Superficies de respuesta estimadas	145
6.10. Autoestima Negativa Retraimiento Modelo no paramétrico	146
6.11. Autoestima Negativa Retraimiento Regresión polinómica	146
6.12. Autoestima Negativa Retraimiento modelo isótono	147
6.13. Proyección sobre el plano Retraimiento-Oposicionamiento Ítem 3	148
6.14. Curvas de nivel	148

8.1. Modelo no paramétrico	168
8.2. Modelo polinómico local	169
8.3. Modelo isótono usando regresión no paramétrica	169
8.4. Modelo isótono usando regresión no paramétrica polinómica	170
8.5. Superficies de respuesta estimadas por regresión NP	171
8.6. Regresión no paramétrica polinómica	172
8.7. Modelo isótono usando regresión no paramétrica	172
8.8. Modelo isótono usando regresión no paramétrica polinómica	173
8.9. Autoestima Negativa Retraimiento Modelo no paramétrico	174
8.10. Autoestima Negativa Retraimiento Regresión polinómica	174
8.11. Autoestima Negativa Retraimiento modelo isótono	174
8.12. Autoestima Negativa Oposicionamiento regresión no paramétrica	174
8.13. Autoestima Negativa Oposicionamiento regresión polinómica	174
8.14. Autoestima Negativa Oposicionamiento modelo isótono	174
8.15. Ineficacia regresión no paramétrica	175
8.16. Ineficacia regresión polinómica	175
8.17. Ineficacia modelo isótono	175
8.18. Ansiedad regresión no paramétrica	175
8.19. Ansiedad regresión polinómica	175
8.20. Ansiedad modelo isótono	175
8.21. Plano Retraimiento-Oposicionamiento (Opción 0)	176
8.22. Plano Retraimiento-Oposicionamiento (Opción 1)	177
8.23. Plano Retraimiento-Oposicionamiento (Opción 2)	178
8.24. Plano Retraimiento-Ineficacia (Opción 0)	178
8.25. Plano Retraimiento-Ineficacia (Opción 1)	179
8.26. Plano Retraimiento-Ineficacia (Opción 2)	179
8.27. Plano Retraimiento-Ansiedad (Opción 0)	180
8.28. Plano Retraimiento-Ansiedad (Opción 1)	180
8.29. Plano Retraimiento-Ansiedad (Opción 2)	181
8.30. Plano Oposicionamiento-Ineficacia (Opción 0)	181
8.31. Plano Oposicionamiento-Ineficacia (Opción 1)	182
8.32. Plano Oposicionamiento-Ineficacia (Opción 2)	182
8.33. Plano Oposicionamiento-Ansiedad (Opción 0)	183
8.34. Plano Oposicionamiento-Ansiedad (Opción 1)	183
8.35. Plano Oposicionamiento-Ansiedad (Opción 2)	184
8.36. Plano Ineficacia-Ansiedad (Opción 0)	184
8.37. Plano Ineficacia-Ansiedad (Opción 1)	185
8.38. Plano Ineficacia-Ansiedad (Opción 2)	185

Índice de cuadros

3.1. Tipos de núcleo	68
3.2. Núcleo beta	69
5.1. Ajuste de los rasgos	116
5.2. Simulaciones seleccionadas	118
5.3. Ajuste en dos dimensiones	118
5.4. Ajuste y tamaño muestral	122
5.5. Ajuste en dos dimensiones. Caso I	122
5.6. Ajuste de las superficies	124
5.7. Ajuste según muestra y cantidad de ítems	125
5.8. Media de los ajustes según ventana	126
5.9. Media de los ajustes modelos logísticos según ventana	126
5.10. Media de los ajustes modelos logit cúbicos según ventana	126
5.11. Ajuste medio por ventana	127
5.12. Ajuste por matriz de ventanas. Caso I	128
5.13. Ajuste por matriz de ventanas. Caso II	128
5.14. Media de los ajustes según núcleo	129
5.15. Media de los ajustes modelos logísticos según núcleo	129
5.16. Media de los ajustes modelos logit cúbicos según núcleo	129
5.17. Ajuste por núcleo caso I	130
5.18. Ajuste por núcleo caso II	131
5.19. Ajuste según distribución	131
5.20. Ajuste por distribución marginal caso I	132
5.21. Ajuste por correlación entre rasgos caso I	132
5.22. Ajuste por distribución marginal caso II	133
5.23. Ajuste por correlación entre rasgos. Caso II	133
5.24. Media de los ajustes según ventana	134
5.25. Media de los ajustes modelos logísticos según ventana	134
5.26. Media de los ajustes modelos logit cúbicos según ventana	134
5.27. Media de los ajustes según núcleo	135
5.28. Media de los ajustes modelos logísticos según núcleo	135
5.29. Media de los ajustes modelos logit cúbicos según núcleo	136

6.1. Ajuste de los cuadernillos 1 y 2.	142
6.2. Parámetros estimados Prueba 1	143
6.3. Parámetros estimados Prueba 2	143

Introducción

La Teoría de Respuesta al Ítem (TRI) es un área de continuo desarrollo dentro de la psicometría y, como reportan Bock (1997) y Thissen & Steinberg (2009), existen cientos de artículos e informes elaborados en las últimas ocho décadas. También ha crecido su participación en investigaciones aplicadas dentro de las ciencias sociales, de la salud y del comportamiento (Reise & Waller, 2009). Actualmente la TRI es una metodología que cumple un rol central en los programas de evaluación alrededor del mundo, como por ejemplo PISA (Program for International Student Assessment) o el NAEP (National Assessment of Educational Progress) en Estados Unidos y el National Assessment of Basic Education Quality en China (Cai et al., 2016). En Uruguay se aplica esta metodología en la prueba adaptativa de inglés a estudiantes de primaria, en la evaluación nacional de aprendizajes llevada a cabo por el INEE (Instituto Nacional de Evaluación Educativa) y en la prueba diagnóstica de competencias al ingreso a la Universidad.

La mayor parte de los artículos y aplicaciones se han realizado para el caso unidimensional de rasgo latente, asumiéndose que la covariación entre los ítems se debe a la presencia de una única variable latente continua, no tomándose en cuenta la presencia de jerarquías anidadas y de variables latentes multidimensionales. Es claro que muchos de los constructos en ciencias sociales y del comportamiento tienen múltiples facetas y rasgos multidimensionales por lo que desarrollar esta parte de la teoría es de capital importancia. Por otro lado, dentro de los modelos estudiados se destacan ampliamente los modelos paramétricos, siendo más limitada la producción académica en modelos no paramétricos. Los modelos paramétricos tienen limitaciones importantes, por ejemplo, no toman en cuenta elementos tales como la no monotonía de los ítems, así como otros apartamientos de la forma propuesta. Otro problema es que los procedimientos de estimación para muchos de estos modelos tienen una fuerte correlación muestral y los procedimientos de estimación son iterativos y la convergencia a la solución es lineal.

Esta tesis se enfoca en el área de los modelos no paramétricos multidimensionales y en modelos no paramétricos con restricciones de monotonía. Estas dos áreas tienen un desarrollo limitado por lo que creemos importante incursionar en ellas.

Las principales contribuciones de esta tesis al estado del arte en los modelos no paramétricos de TRI son:

- Presentar un modelo unidimensional de TRI basado en regresión polinómica local no paramétrica.
- Plantear un estimador de los rasgos y un estimador de las CCI basado en regresión no paramétrica para el caso multidimensional, generalizando así el modelo unidimensional de Ramsay y el basado en regresión mediante polinomios locales. Estos están dados por las ecuaciones 3.24, 3.7 y 3.11.
- Contemplar en los modelos propuestos las variantes para ítems dicotómicos, politómicos y respuesta continua (sección 3.2.2).

- Presentar estimadores de la densidad del rasgo y de la cópula subyacente (ecuaciones 3.14 y 3.16).
- Dar propiedades de los estimadores cuando el rasgo es medido sin error. Probar la consistencia en la estimación de la densidad del rasgo y su distribución asintótica (teoremas 3.3.2, 3.3.9), 3.3.8 y subsección 3.3.1).
- Dar aproximaciones a la elección de la ventana óptima para la estimación de la densidad (sección 3.5).
- Establecer las condiciones que deben cumplirse para que los estimadores multidimensionales del rasgo y la CCI propuestos sean consistentes de manera conjunta en el caso general.

Hipótesis: sección 3.7.

Consistencia de los rasgos cuando no hay empates: teorema 3.8.10.

Consistencia de los rasgos una vez rotos los empates: corolario 3.8.12.

Cota del ECM de las CCI: teorema 3.8.14.

Consistencia de las CCI: teorema 3.8.15.

Consistencia conjunta: teorema 3.8.16.

- Probar la consistencia de la estimación de la CCI, su distribución asintótica y la convergencia completa en el caso donde el rasgo es medido sin error (teoremas 3.9.1, 3.9.4, 3.9.5, 3.9.6, 3.9.7 y 3.9.8).
- Dar aproximaciones al sesgo y la varianza y a la elección de la ventana óptima para la estimación de la CCI (secciones 3.10 y 3.10.1).
- Plantear un nuevo estimador monótono de las CCI para el caso unidimensional (sección 4.1).
- Encontrar un estimador monótono para las CCI en cada componente para el caso multidimensional (sección 4.1).
- Dar propiedades para el estimador isotónico en el caso de rasgo medido sin error, (teorema 4.2.13).
- Realizar simulaciones para determinar el comportamiento de los estimadores hallados (capítulo 5).
- Aplicar estos modelos a tests específicos para estudiar su utilidad práctica (capítulo 6).

La tesis se estructura en siete capítulos y un anexo que abarcan aspectos teóricos, simulaciones y aplicación de los modelos hallados a casos reales. Los capítulos I y II presentan el estado actual de las investigaciones en TRI y TRI no paramétrico. En los capítulos III al VI se desarrollan los aportes específicos de la tesis y, finalmente, el capítulo VII trata de las conclusiones obtenidas, las limitaciones de los métodos y los problemas abiertos que surgen a partir de lo estudiado. En el anexo se presentan los gráficos de modo de facilitar la lectura de la tesis.

En el capítulo I se realiza una brevísima presentación sobre TRI paramétrico. Se consideró importante analizar con cierto detalle los orígenes de la TRI y sus supuestos y también exponer claramente algunos métodos de estimación y sus algoritmos. El objetivo de esto último es

evidenciar las dificultades en la estimación de los parámetros y los rasgos en este tipo de modelos para contrastarlos luego con la simplicidad de los algoritmos de estimación no paramétricos. Específicamente en este capítulo nos focalizamos en cuatro aspectos:

a) Orígenes, primeros desarrollos y supuestos de la TRI, presentadas desde una perspectiva que no es la que usualmente está presente en la literatura.

b) Una somera descripción de los modelos paramétricos que pueden analizarse también desde una perspectiva no paramétrica.

c) Descripción detallada de los algoritmos de estimación para los modelos clásicos a los efectos de mostrar su complejidad, que la solución obtenida puede ser a menudo inestable y la indeterminación de la escala, de forma de poder compararlos con la simplicidad de los algoritmos propuestos.

d) Presentación de algunos modelos multidimensionales paramétricos y sus supuestos, ya que esta tesis tiene un alto componente en este tipo de modelos desde la perspectiva no paramétrica.

En el capítulo II se presentan los métodos no paramétricos y en particular el modelo de Ramsay unidimensional como antecedente de los discutido en esta tesis, así como un nuevo enfoque de Ramsay del año 2016.

El capítulo III está dedicado a la presentación de los dos modelos multidimensionales no paramétricos mencionados anteriormente. Se desarrollarán los siguientes temas:

a) Estimación del rasgo.

b) Método para la estimación aproximada de la densidad del rasgo, basado en la densidad no paramétrica para variables aleatorias observables (rasgo medido sin error).

c) Estimación de las CCI y FRI para los diferentes casos.

d) Consistencia del estimador y distribución asintótica en el caso de rasgos medidos sin error.

e) Una sección, a nuestro juicio muy importante, dedicada a encontrar las condiciones para que el estimador multidimensional basado en regresión no paramétrica para el caso dicotómico sea consistente conjuntamente para el rasgo y las CCI, esto es cuando el tamaño muestral y la longitud del test crecen juntos. Las condiciones encontradas tratan sobre el núcleo, los anchos de las ventanas y las funciones auxiliares usadas en la estimación de los rasgos. Bajo estas condiciones se obtienen cuatro resultados principales en R^d que incluyen el resultado de Douglas (1997) para el caso unidimensional:

a) Consistencia del rasgo: $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad y

$\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente (en particular con probabilidad 1).

b) Para $\theta_0 \in (0, 1)^d$ fijo

$$(1) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

c) Consistencia de las CCI: $\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$ si $n \rightarrow +\infty$ completamente con $(a,b)^d \subset (0, 1)^d$.

d) Consistencia conjunta: con probabilidad 1.

$$\max\{\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0 \text{ si } n \rightarrow +\infty$$

En el capítulo IV se presenta el modelo isótono unidimensional y el modelo isótono en cada dimensión para el caso multidimensional. Nos basaremos en un método desarrollado por Dette et al. (2006). A partir del estimador no monótono de la CCI se obtiene el estimador isótono de la inversa de la CCI mediante

$$(2) \quad \widehat{P^{-1}}(\theta) = \frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T}) - u}{h_d}\right) du$$

luego, a partir de esta, obtenemos la CCI. Para el caso de rasgo conocido o medido sin error siguiendo el trabajo de Dette et al.(2006) mostramos como quedaría la distribución asintótica del estimador.

En el capítulo V se muestran los resultados de los estudios de simulación llevados a cabo para analizar diferentes aspectos de los modelos presentados.

En el capítulo VI aplicamos los modelos a dos casos particulares:

a) Analizamos dos cuadernillos de la prueba de evaluación de conocimientos en matemática al ingreso a la Universidad de la República en Uruguay, donde aplicamos los cuatro modelos unidimensionales no paramétricos presentados y luego aplicamos los modelos al caso bidimensional, considerando las dimensiones álgebra y cálculo.

b) Estudiamos el Children Depression Inventory (CDI) de Kovacs, aplicado a una muestra representativa de estudiantes de Montevideo. Este test es de respuestas politómicas por lo que usamos los modelos no paramétricos presentados en el caso de respuesta graduada. Luego, estimamos las funciones de respuesta considerando un modelo de dimensión cuatro y graficando las proyecciones en un espacio de dimensión dos.

Finalmente, en el capítulo VII presentamos las conclusiones de la tesis, así como las limitaciones de los modelos. También mencionamos posibles líneas de desarrollo futuro y problemas abiertos.

Capítulo 1

Teoría de respuesta al ítem

1.1. Orígenes y primeros desarrollos

En este capítulo presentamos en forma sintética los principales modelos paramétricos de la teoría de respuesta al ítem (TRI) que se han desarrollado, dejando la presentación de los modelos no paramétricos para el próximo capítulo. Como explicamos en la introducción, solo presentaremos con detalle los aspectos directamente relacionados con la tesis y en forma breve los otros modelos a los efectos de realizar una síntesis histórica de la TRI desde sus inicios a nuestros días. La única excepción a esto se dará al inicio de este capítulo donde trataremos los orígenes y supuestos de la TRI pretendiéndose con esto presentar algunos elementos que usualmente no son mencionados en la bibliografía usual. Dedicaremos atención a los modelos que pueden ser analizados desde la perspectiva no paramétrica basada en regresión con núcleos, tanto en una dimensión como en dimensiones superiores. También nos focalizaremos, en parte de este capítulo, en presentar los algoritmos de estimación de los modelos clásicos, su deducción e implementación, para poder justificar luego nuestra afirmación que los métodos propuestos son mucho más simples de estimar que los paramétricos usuales y con performance similar respecto al ajuste cuando los datos siguen el modelo paramétrico propuesto y mucho mejor cuando los datos no lo siguen.

La TRI se diferencia de la teoría clásica de los test por utilizar modelos basados en las características de los ítems en vez de las del test, donde estas características son independientes del grupo en que el ítem se ha calibrado y las puntuaciones en el rasgo no dependen de las puntuaciones obtenidas en cada test particular. Otra diferencia importante reside en el modelo inicial del que se parte, que permite obtener una medida de la precisión para cada puntuación del rasgo, lo que la distingue claramente de la teoría clásica y también permite que para evaluar la fiabilidad no se requieren de tests estrictamente paralelos. La TRI toma los ítems como unidad de análisis y permite describir algunas propiedades psicométricas del test mediante indicadores invariantes. Este enfoque establece una relación funcional entre la respuesta del examinado a cada ítem y el rasgo latente responsable de tal realización y al que notaremos θ . La función que valora la probabilidad de obtener determinada puntuación en el ítem condicionado al rasgo en los modelos dicotómicos se denomina usualmente curva característica del ítem (CCI) y en el caso politómico funciones de respuesta al ítem (FRI). En esta tesis se utilizará cualquiera de los términos como sinónimos siempre que no dé lugar a confusión y en el caso multidimensional se denominarán superficies de respuesta al ítem.

Se ha adjudicado la primera referencia del concepto *curva característica del ítem* a Lord y Novick (1968). Sin embargo, Carlson y von Davier (2013) rescatan una cita de Green (1980)

donde se señala a Tucker (1946) como quien introdujo este concepto usando el modelo de ojiva normal. Efectivamente, Tucker en su artículo *Maximum validity of a test with equivalent items* publicado en *Psychometrika* en 1946 presenta una serie de ideas básicas para el posterior desarrollo de la TRI. En primer lugar, asume que existe una función sobre los verdaderos valores de una escala y que la probabilidad de responder correctamente un ítem es una curva dada por la integral de la distribución normal. También, esboza el concepto de variable latente continua, al afirmar que existe un continuo de puntajes verdaderos, los que con los errores de medida, producen los puntajes observados. Presenta gráficamente el modelo de ojiva normal y expresa : “se asume que los individuos con puntajes bajos en el continuo de la habilidad son menos propensos a responder correctamente cada ítem e individuos con altos puntajes verdaderos tienen más probabilidad de responder cada ítem correctamente...La probabilidad de éxito en un ítem estaría relacionada con los puntajes verdaderos en la habilidad...Un ítem que tiene discriminación perfecta estará representado por una línea vertical localizada en aquel valor de la habilidad que todos los individuos con puntaje menor fallan el ítem y todos los individuos con puntaje más alto lo pasan” (Tucker, 1946: 2).

Más adelante introduce el modelo de ojiva normal de dos parámetros y establece que la “proposición precedente puede establecerse más precisamente en términos matemáticos. Las curvas de los ítems...en el caso más simple depende de dos parámetros: uno, para el nivel general de dificultad, y otro, para la potencia discriminativa del ítem” (Ibidem: 2).

En la ecuación 2 del artículo mencionado Tucker (1946) representa la curva mediante:

$$p_{js} = \int_{u_j=-\infty}^{u_j=(s-s_j)/\sigma_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu_j^2} d\mu_j$$

siendo s_j es el valor de la habilidad donde la probabilidad de acertar el ítem es 1/2 y σ_j es una constante de las curvas que determina la dispersión de la curva.

En la década de 1950 Fred Lord y Bert Green realizaron importantes aportes a la teoría de los tests de forma que sus trabajos se pueden considerar algunos de los primeros aportes a la TRI. Es interesante observar como en esa época los boletines del Educational Testing Service eran sumamente detallados y dieron origen a diferentes artículos. Ejemplos de estos boletines de investigación son el Research Bulletin No. RB-50-38 de Green (1950) (“*A general solution for the latent class model of latent structure analysis*”) y el Research Bulletin No. RB-52-10 de Lord (1952)(“*The relation of test score to the ability underlying the test*”). Estos reportes condujeron a sendos artículos, Green (1951a) publicado en *Psychometrika* y Lord (1953) publicado en *Educational and Psychological Measurement*.

El trabajo de Lord y Green en esos años fue muy fructífero, Lord (1951,1952a, 1952b, 1953) usó y explicó muchos de los conceptos de la TRI como curva característica del ítem, curva característica del test, error estándar condicionado a la habilidad latente y el concepto de invarianza de los parámetros. Aparece en esa época el concepto de *independencia local* que Lord (1980a) atribuye a Lazarsfeld (1950) quien lo usa en el capítulo “*The logical and mathematical foundation of latent structure analysis*”. del libro *Studies in social psychology in World War II: Vol. 4. Measurement and prediction*. Lord (1980a) menciona además que Lazarsfeld usó el término *trace line* para una curva como la CCI, entendiendo como curva característica del ítem a la relación funcional que expresa la probabilidad de responder correctamente a un ítem j condicionado a una variable latente.

Green, por su parte, en 1950, presenta en el Educational Testing Service el reporte de investigación RB-50-38 que ya mencionamos, y además, el RM-50-20 (“*A proposal for a comparative study of the measurement of attitude*”), RM-50-26 (“*A proposal for an empirical evaluation of the latent class model of latent structure analysis*”) y RB-50-65 (“*Latent structure analysis and*

its relation to factor analysis”) y en 1951 el RB-51-15 (“Latent class analysis: A general solution and an empirical evaluation”) y el artículo en *Psychometrika* también ya mencionado donde analizó datos de respuesta al ítem usando estructuras latentes y modelos de clase latente (Green 1950a, b, c, d). En 1952 publica un artículo en el *Journal of the American Statistical Association* (“Latent structure analysis and its relation to factor analysis”) donde mostró la relación entre los modelos de estructura latente y el análisis factorial.

El modelo de ojiva normal de Lord

Este es el primer modelo desarrollado y utiliza desde una nueva perspectiva los dos parámetros clásicos usados por la teoría clásica de los tests: la discriminación y la dificultad de los ítems.

Lord lo presenta en 1952 en su publicación “A theory of test scores ” obteniendo su deducción de una manera natural y donde resulta magnífica la adecuación de elementos de la psicofísica al problema de respuesta de los ítems considerando una variable latente que condiciona las respuestas. Sin embargo, el modelo presentaba dificultades para la estimación de los parámetros, que fueron solucionadas más tarde por Lord y Novick en 1968.

Lord considera una variable auxiliar $\Gamma_i(\theta)$ (la cual depende del rasgo), que indica la propensión a responder correctamente el ítem i en función del rasgo, con $-\infty \leq \Gamma_i \leq \infty$, y luego asume la existencia de un umbral γ_i tal que si $\Gamma_i \geq \gamma_i$, el ítem se responde correctamente, es decir, $u_{ij} = 1$ y si $\Gamma_i < \gamma_i$ el ítem se falla, o sea $u_{ij} = 0$. Claramente la variable Γ_i no es observable, siendo observables las respuestas de los sujetos a cada ítem, esto es u_{ij} con $j = 1, 2, \dots, N$. Lord asume para modelizar a los ítems que: Γ_i es una función lineal de θ , es decir, $\Gamma_i = \alpha_i + \beta_i \theta + \epsilon_i$ con $\epsilon_i = N(0, \sigma^2)$ y donde el rasgo se considera sin errores de medida. Deduce entonces la probabilidad de responder correctamente el ítem i dado que el sujeto tiene un nivel de rasgo θ_j mediante:

$$(1.1) \quad P_i(\theta_j) = \int_{-\infty}^{a_i(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Al parámetro b_i le llama dificultad del ítem y representa el punto en la escala de rasgo en el cual la probabilidad de responder correctamente el ítem es 0.5. Al parámetro a_i le llama índice de discriminación del ítem i y es proporcional a la pendiente de la CCI en el punto $\theta = b_i$, como se puede comprobar directamente derivando y observando que $P'_i(b_i) = \frac{a_i}{\sqrt{2\pi}}$.

En la deducción del modelo, Lord no hace supuestos sobre la distribución del rasgo en la población. En Baker (1992) se puede ver la deducción completa del modelo y sus propiedades.

Si la distribución es normal y no hay aciertos al azar en el ítem, se puede establecer una conexión con la teoría clásica de los test. Richardson (1936) y Tucker (1946) mostraron que bajo esta condición el coeficiente de correlación biserial $\rho_{\theta U_i}$ cumple $\rho_{\theta U_i} = \frac{a_i}{\sqrt{1+a_i^2}}$. Además, si notamos la dificultad del ítem (según la teoría clásica) por p_i y notando $\delta_i = b_i \rho_{\theta U_i}$; Tucker (1946) mostró que:

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{\delta_i}^{+\infty} e^{-\frac{1}{2}z^2} dz$$

El modelo que acabamos de presentar es conocido como modelo de ojiva normal de dos parámetros. Existen variaciones de este modelo. Si suponemos que la discriminación es uno, se obtiene el modelo de un parámetro. Birnbaum (1968) incorpora la posibilidad de aciertos al azar agregando un parámetro y deduce el modelo de tres parámetros que toma la forma $P_i(\theta_j) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta_j - b_i)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$, en tanto Barton y Lord (1981) han propuesto un modelo de cuatro parámetros, que no discutiremos aquí, ya que no aporta ventajas significativas respecto a los anteriores.

Sijstma y Junker (2006) hacen notar que el modelo de Lord es básicamente el modelo de juicio comparativo de Thurstone (1927). El modelo de juicio comparativo expresa la probabilidad que un estímulo j sea preferido sobre un estímulo k usando la función de distribución normal mediante:

$$P(t_{jk} > 0) = P(\epsilon_{jk} > -(\mu_j - \mu_k)) = \int_{-(\mu_j - \mu_k)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

donde μ_j y μ_k son las medias de los estímulos, ϵ_{jk} un error con distribución normal típica y $t_{jk} = \mu_j - \mu_k + \epsilon_{jk}$. Si en lugar de comparar dos estímulos se realiza la comparación de una persona v y un ítem j el problema será analizar si la persona v domina el ítem j o sea si $(\theta_v > b_j)$. Para esto asumamos un error con distribución normal típica, ϵ_{vj} , definamos $t_{vj} = a_j(\theta_v - b_j) + \epsilon_{vj}$, y entonces, usando la simetría de la distribución normal, el modelo que resulta es:

$$P_j(\theta) = P(t_{vj} > 0) = P[\epsilon_{vj} > -a_j(\theta_v - b_j)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_j(\theta_v - b_j)} e^{-t^2/2} dt; a_j > 0$$

Observemos que la diferencia sustancial entre los dos modelos es que el de Lord incluye una variable latente en su formulación y que esta variable está en la misma dimensión que mide a los ítems; también permite a las curvas de respuesta, pendientes que varían entre ítems.

Modelos determinísticos

Hay dos modelos determinísticos, el modelo de Guttman y el modelo de Coombs que pueden considerarse también dentro de los precursores de la TRI (Sijstma y Junker, 2006). El modelo de Guttman (1944 y 1950) compara los parámetros de localización de los ítems y de las habilidades de las personas, pero sin hacer intervenir un modelo probabilista. Si b_i es el parámetro de localización del ítem i , entonces el modelo de Guttman establece que la curva de respuesta al ítem es una función escalonada con un salto de una unidad en b_i . Es decir, el modelo de Guttman predice con certeza el valor del ítem de acuerdo a que $\theta < b_i$ o $\theta \geq b_i$; si $\theta < b_i$ el valor del ítem es 0, si $\theta \geq b_i$ el valor del ítem es 1. Concretamente, $\theta < b_i \Leftrightarrow P_i(\theta) = 0$; y $\theta \geq b_i \Leftrightarrow P_i(\theta) = 1$. Expresado como una función del signo de $(\theta - b_i)$, siendo χ la indicatriz de un conjunto, el modelo será: $P_i(\theta) = \chi_{[b_i, +\infty)}(\theta)$. Este tema lo retomaremos más adelante cuando hablemos sobre TRI no paramétrico, ya que Mokken (1971) toma como base este modelo para el desarrollo de su modelo no paramétrico. También Coombs (1964) desarrolló un modelo determinístico, que es similar al de Guttman en tanto predice con certeza la puntuación en el ítem. Guttman utiliza el parámetro de localización b_i y un parámetro d_i llamado "*latitud de aceptación*". En este modelo el sujeto responde correctamente al ítem si su rasgo está en un entorno (bola) cerrado centrado en b_i y de radio $d_i/2$.

$$P(U_i = 1 | \theta) = \begin{cases} 1 & \text{si } |\theta - b_i| \leq d_i/2 \\ 0 & \text{en otro caso} \end{cases}$$

Un aporte particularmente importante de Coombs (1964) es la utilización del concepto de *desdoblamiento*, descrito como una metáfora geométrica para reconciliar el conflicto del orden de preferencias dado por las diferentes respuestas a un conjunto de estímulos, que se utilizó más tarde para el desarrollo de los modelos de desdoblamiento, los modelos de proximidad, los modelos de punto ideal y los modelos unimodales de respuesta al ítem.

Curva característica del ítem

Si consideramos un ítem dicotómico i su distribución de frecuencias viene dada por:

$f_i(u/\theta) = P_i(\theta)^u Q_i(\theta)^{1-u}$ donde $u = 1$ o 0 y $Q_i(\theta) = 1 - P_i(\theta)$. La función $P_i(\theta)$ es la esperanza condicional de la puntuación del ítem dado el nivel de rasgo, es decir, es la regresión de la puntuación del ítem sobre la aptitud. A esta función se le llama función o curva característica del ítem (CCI), si estamos en el caso de modelos dicotómicos.

Definición 1.1.1. Llamaremos curva (superficie en el caso mutidimensional) característica del ítem a la función: $P(\theta) : A \subset R^d \mapsto [0, 1]$ que da la probabilidad de acertar el ítem condicionada a la variable latente θ , es decir,

$$(1.2) \quad P(\theta) = P(U = 1/\theta)$$

Si el espacio latente está completamente especificado, las CCI se mantendrán invariantes para todo subgrupo de la población; luego la probabilidad que un individuo dé una respuesta correcta a determinado ítem solo depende de la forma de la curva característica del ítem y del nivel de rasgo del sujeto, y entonces, es independiente de la distribución del rasgo en la población. Esta propiedad de invarianza de las CCI (no importa en qué región fue calibrado el ítem) constituye una de las mejores características de los modelos de TRI y es de capital importancia en las aplicaciones. La formulación matemática de las CCI es lo que distingue unos modelos de otros.

De manera análoga, se puede hablar de curva característica del test (CCT), que se define como la suma de las curvas características de los ítems que componen el test, o sea, $CCT = \sum_{i=1}^n P_i(\theta)$.

También se puede estudiar la curva característica del sujeto que relaciona la dificultad del ítem y la proporción de ítems respondidos correctamente para cada nivel de dificultad. Esta es una buena herramienta para comparar el proceder de las diferentes personas a través de los ítems, así como para estudiar las curvas empíricas de las personas y las que cabría esperar si se ajustasen a las predicciones del modelo.

1.2. Supuestos de la teoría de respuesta al ítem

Mencionaremos los tres supuestos básicos que están en el inicio de los modelos de respuesta al ítem y que aún hoy se mantienen, aunque existen modelos que violan alguno de ellos.

Dimensionalidad del espacio latente

Se supone que hay un conjunto de rasgos o aptitudes que subyacen en la respuesta de los individuos a un conjunto de ítems. Cada rasgo es un número real, por lo tanto, si existen d rasgos estos determinan puntos del espacio R^d (o un subconjunto de él). A los modelos que suponen que un único rasgo es el que determina las respuestas se les llama unidimensionales. Es claro que este supuesto nunca se cumplirá estrictamente, por lo tanto, lo que se pide es que para un conjunto de ítems del test exista un factor dominante que explique la realización del test, a este factor se le llama aptitud medida por el test. Los modelos que consideran más de un factor son llamados multidimensionales. Si se cumple el supuesto de unidimensionalidad, al aplicar a r subpoblaciones de examinados un test, las distribuciones condicionadas de las puntuaciones del test para cada nivel de rasgo son iguales para todas las subpoblaciones. En caso de no cumplirse lo anterior estamos en presencia de funcionamiento diferencial del ítem.

A los efectos de la construcción de tests unidimensionales, Lumdsen (1961) recomienda el uso del análisis factorial. A un conjunto inicial de ítems (construidos a partir de conocimientos previos) se les realiza un análisis factorial y se eliminan aquellos que no carguen en el primer factor. Una vez eliminados esos ítems se repite el procedimiento tantas veces como sea necesario, hasta dar con un solución satisfactoria. También propone que la razón de varianza explicada

por el primer factor al segundo se utilice como índice de unidimensionalidad. No entraremos en detalle en este aspecto, solo mencionamos que se han desarrollado numerosos métodos que se utilizan para determinar la dimensionalidad de un test.

Independencia local

La independencia local refiere a que los n ítems que constituyen el test, condicionados a un valor constante de $\theta \in R^d$ son independientes. Consideremos un test formado por n ítems y sean U_1, U_2, \dots, U_n las variables que representan las respuestas de los examinados a los ítems. Por ejemplo, si los ítems son dicotómicos, las U_i son variables Bernoulli (acierto, error) que indican el resultado en cada ítem. Entonces, si hay independencia local se cumple: $P(U_1 \dots U_n / \theta) = \prod_{j=1}^n P(U_j / \theta)$.

La condición anterior es necesaria, pero no suficiente para la independencia dos a dos de las variables. Una alternativa posible para una definición formal de la independencia local con una condición necesaria y suficiente puede ser:

Definición 1.2.1. n ítems son localmente independientes condicionados al rasgo θ , si para cualquier subconjunto de índices $i(1), \dots, i(k)$, $1 \leq k \leq n$ $1 \leq i(j) \leq n$

$$(1.3) \quad P(U_{i(1)} \dots U_{i(k)} / \theta) = \prod_{j=i(1)}^{i(k)} P(U_j / \theta)$$

Es importante notar que la independencia local no implica que los ítems son no correlacionados sino que son independientes condicionalmente (al nivel de rasgo). Hambleton y Swaminathan (1991) muestran que el principio de unidimensionalidad es equivalente a la independencia local (condicionado a solo un rasgo) pues si suponemos que el test es unidimensional y mide una variable latente θ , si no se da la independencia local resultaría que algunos sujetos tendrían mayor probabilidad de responder correctamente para el mismo nivel de rasgo, lo cual niega la unidimensionalidad. Recíprocamente, si se da la independencia local tenemos que las probabilidades de respuesta solo dependen de un único rasgo. La independencia local también puede expresarse con datos multidimensionales (condicionando al espacio completo).

Dada la equivalencia entre independencia local y unidimensionalidad, cuando se consideran datos unifactoriales, McDonald (1981) propone usar técnicas de análisis factorial para evaluar el supuesto de independencia local. Define un conjunto de ítems del test como unidimensionales, cuando para sujetos con el mismo nivel de rasgo, la covariación entre los ítems es cero. Molenaar y Sijtsma (2000) usan técnicas de cluster para encontrar la dimensionalidad del test.

La ventaja de considerar los ítems localmente independiente resulta obvia a la hora de querer estimar los parámetros de los ítems dado que la verosimilitud se convierte en un producto de probabilidades referidas a la respuesta de un solo sujeto. Implica también que la habilidad es suficiente para poder determinar la probabilidad de respuesta de un sujeto a un ítem k , independientemente de los resultados obtenidos en los otros ítems. Para ver esto consideremos que se aplica a un sujeto un test de n ítems y sea $U_{-k} = (U_1, U_2, \dots, U_{k-1}, U_{k+1}, \dots, U_n)$ el vector de respuestas sin considerar el ítem k . Entonces:

$$(1.4) \quad P(U_k / \theta, U_{-k}) = \frac{P(U_1 \dots U_n / \theta)}{P(U_{-k} / \theta)} = \frac{\prod_{j=1}^n P(U_j / \theta)}{\prod_{j=1, j \neq k}^n P(U_j / \theta)} = P(U_k / \theta)$$

Esta ecuación nos indica que el proceso de responder no influye en el patrón de respuestas, es decir, cada ítem se modeliza independientemente de los demás. Es decir, el modelo completo se compone por salidas independientes y la probabilidad conjunta será el producto de las probabilidades de las respuestas a cada ítem.

No siempre se cumple la independencia local, se puede hablar de dependencia local, por ejemplo, en el caso en que exista un entrenamiento de los sujetos. El modelo de Rasch multidimensional de Embretson (1991) para aprendizaje y cambio trata esta cuestión y Jannarone (1997) presenta modelos para ítems localmente dependientes. Bradlow, Wainer y Wang (1999) dan modelos para testest donde la dependencia viene dada al depender conjuntos de ítems de un test de un estímulo común.

Si observamos la ecuación 1.4 veremos que la independencia local es una condición exigente, ya que al suponer que la respuesta a un ítem dado no depende del patrón de respuestas en los otros ítems no permite, por ejemplo, el aprendizaje durante la administración del test. Si esto llega a ocurrir, entonces, la independencia local es violada; o en otro caso, por ejemplo, al realizar un test de opinión los sucesivos ítems pueden llevar a reflexionar más profundamente sobre lo que se contesta y por lo tanto influir en los ítems sucesivos. Se han propuesto algunas alternativas a la independencia local para debilitar el supuesto.

Definición 1.2.2. Independencia Esencial (Stout 1990). Los ítems serán esencialmente independientes si $\lim_{n \rightarrow \infty} \frac{\sum_{1 \leq j < k \leq n} |Cov(U_j, U_k/\theta)|}{\binom{n}{2}} = 0$

Definición 1.2.3. Independencia Local Débil (Zhang y Stout 1999a). Los ítems serán localmente independientes débilmente si: $Cov(U_j, U_k/\theta) = 0$. Es claro que la independencia local implica la independencia local débil.

Observaciones sobre el supuesto de monotonía

Como señalan Sijstma y Junker (2006) la monotonía es un supuesto fuerte y podemos afirmar, en función de lo que desarrollaremos en este apartado, que también es discutible. La argumentación a favor de la no monotonía es compleja de explicar; las aplicaciones en datos reales muchas veces aportan en este sentido. Bouwmeester, Sijstma y Vermunt (2004) usan regresión de clases latentes para describir el desarrollo cognitivo y dan una explicación basada en las diferentes estrategias de resolución de los ítems que los sujetos situados en diferentes niveles de habilidad pueden tener. Estas estrategias varían en complejidad y en su adecuación al problema. Esto puede provocar en algunos niveles de rasgo desviaciones de la monotonía. Este supuesto se torna particularmente fuerte cuando se debe mantener para todos los n ítems del test (Sijstma y Junker, 2006). Por eso, Stout (1987,1990) propuso para ítems dicotómicos replazar la monotonía por la monotonía débil.

Definición 1.2.4. Monotonía Débil. Se cumple la monotonía débil si la función de respuesta del test: $T(\theta) = \frac{\sum_{j=1}^n P_j(\theta)}{n}$ es creciente con respecto a cada coordenada de θ , para n suficientemente grande.

La monotonía débil garantiza que hay suficiente información para estimar θ a partir de los puntajes en el test. Stout (1990) mostró que si conjuntamente la monotonía débil y la condición de independencia esencial se cumplen en el caso de θ unidimensional, entonces el puntaje total $X_+ = \sum_{j=1}^n U_j$ estima consistentemente θ cuando $n \rightarrow \infty$; resultado que fue generalizado para ítems politómicos por Junker (1991). Si la monotonía se cumple para todos los ítems, y se mantienen la independencia local y la unidimensionalidad, entonces si $x_{+v} < x_{+w}$, para cada t , tenemos que $P(\theta > t \mid X_+ = x_{+v}) \leq P(\theta > t \mid X_+ = x_{+w})$. Es decir, se puede establecer un orden estocástico entre los sujetos y esto implica que $E(\theta \mid X_+)$ es monótona creciente como función del puntaje total X_+ .

Estos resultados incluyen casos especiales como el modelo de Rasch (1960) y el modelo logístico de 3P, al igual que el modelo de ojiva normal de Lord. Hemker et al. (1997) mostraron que lo anterior también se mantiene para el modelo de crédito parcial para ítems politómicos ordenados, pero no para otros modelos politómicos. Van der Ark (2005) estudia las consecuencias prácticas y la robustez del orden estocástico en varios modelos politómicos.

La estimación de la función de respuesta del ítem para un conjunto de datos reales tiende a ser monótona, pero la no monotonía local es usual. Algunos análisis asumen la monotonía débil y hasta el supuesto de pérdida de monotonía porque en la práctica n es finito y a menudo pequeño y en ese caso la función de respuesta del ítem no monótona puede conducir a serias distorsiones en el ordenamiento de personas por X_+ . Sin embargo, solamente permitir ítems que satisfagan la monotonía en un test parece ser demasiado estricto, porque la no monotonía es bastante frecuente.

1.3. Modelos para ítems dicotómicos

Los ítems dicotómicos son aquellos en los que una respuesta de un sujeto será clasificada en alguna de dos categorías, según responda correctamente o incorrectamente. Se puntuará el ítem 1 si se responde correctamente y 0 si se hace incorrectamente. Si un individuo no responde a un ítem se considerará su respuesta como incorrecta. Muchos modelos de TRI han sido definidos para puntajes de ítems dicotómicos en base a los supuestos de independencia local y unidimensionalidad a la que se le agregó la versión paramétrica del supuesto de monotonía. Hemos ya visto el modelo de ojiva normal de Lord en sus distintas versiones y hemos mencionado las dificultades iniciales de su estimación. Así, debido a los méritos computacionales, aparecieron los modelos logísticos los que han ganado gran popularidad, también debido a su simpleza con respecto a la integral que aparece en el modelo de ojiva normal. Otro motivo, a mi criterio, es que estos modelos se relacionan con modelizar el logit linealmente (logaritmo de la razón de probabilidades), y la razón de probabilidades es un concepto usual en la cultura anglosajona donde se desarrollaron inicialmente estos modelos.

Consideremos la función $\Psi(x) = \frac{e^x}{1+e^x}$ y sea $\Phi(x)$ la distribución normal acumulada, Haley (1952) mostró que $|\Phi(a_i(\theta_j - b_i)) - \Psi((1.702)a_i(\theta_j - b_i))| < .01$. A partir de esta aproximación, utilizándola en los modelos de ojiva normal, Birbaum (1968) obtiene los modelos logísticos correspondientes (donde $D=1.702$):

$$(1.5) \quad 1P - P_i(\theta_j) = \frac{e^{D(\theta_j - b_i)}}{1 + e^{D(\theta_j - b_i)}}$$

$$(1.6) \quad 2P - P_i(\theta_j) = \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}$$

$$(1.7) \quad 3P - P_i(\theta_j) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}$$

En general, el parámetro D se incorpora a los otros y, por lo tanto, de ahora en adelante supondremos que $D=1$. Como ya hemos visto para los modelos de ojiva normal tenemos que a_i es el parámetro de discriminación, b_i el parámetro de dificultad y c_i el parámetro que indica la posibilidad de aciertos al azar. Estos modelos tienen la ventaja de ser mucho más manejables matemáticamente y son los que comúnmente se usan en TRI. El modelo de 3P es apropiado en particular para ajustar datos de ítems de múltiple opción con dificultad y discriminación variada y cuando es probable que los examinados con baja habilidad contesten al azar. Se han hecho intentos para agregar un asíntota superior al modelo de 3P a los efectos de hacerlos más

robustos (Barton y Lord, 1981), pero este modelo de 4 parámetros no ha resultado adecuado. Un mejor modelo es el de Rasch agregando un parámetro de adivinación (San Martín et al. 2006).

Bock y Lieberman (1970) reformularon los modelos de efectos fijos de 3P de ojiva normal y logístico como un modelo donde los sujetos son efectos aleatorios, es decir, son muestreados independientemente de una población, por lo tanto, hay una dependencia de la función de distribución del rasgo. Con esto se logran varias ventajas, como señala De Boeck (2008). Una aplicación natural de la TRI con efectos aleatorios es la modelización de las distribuciones de los parámetros de los ítems en diferentes familias de ítems generados por distintas condiciones en un generador de ítems basado en reglas.

Para lograr un mayor significado del modelo en su interpretación y para relacionarlo con las regresiones logísticas, se pueden utilizar los logits, definidos como: $\ln \frac{P(u_{ij}=1|\theta)}{P(u_{ij}=0|\theta)}$. Se verifica que los logits para los diferentes modelos son: $\theta - b_i$ (modelo 1P); $a_i(\theta - b_i)$ (modelo 2P) y $a_i(\theta - b_i)$ (tomando $\ln \frac{P(u_{ij}=1|\theta)-c_i}{P(u_{ij}=0|\theta)}$) para el modelo 3P. El logit también es útil para comparar sujetos. pues:

$$\ln \left[\frac{\frac{P(u_{ij}=1|\theta_1)}{P(u_{ij}=0|\theta_1)}}{\frac{P(u_{ij}=1|\theta_2)}{P(u_{ij}=0|\theta_2)}} \right] = \ln \frac{P(u_{ij}=1|\theta_1)}{P(u_{ij}=0|\theta_1)} - \ln \frac{P(u_{ij}=1|\theta_2)}{P(u_{ij}=0|\theta_2)}$$

Esta diferencia según el caso es igual a $\theta_1 - \theta_2$ para el modelo 1P y $a_i(\theta_1 - \theta_2)$ para los modelos 2P y 3P. Es claro que los modelos logísticos se pueden también obtener del hecho de suponer que los logits son una función lineal de la habilidad. Por ejemplo, para el modelo de dos parámetros,

$$(1.8) \quad \ln \frac{P(u_{ij}=1|\Theta=\theta_j)}{P(u_{ij}=0|\Theta=\theta_j)} = a_i(\theta_j - b_i).$$

Identificabilidad de la escala de θ

Un problema no menor dentro de la TRI es la identificabilidad de la escala. Un hecho importante es que un test no puede dar más información que sobre la ordenación de los examinados. Este hecho lo utilizaremos reiteradamente cuando elaborem y presentemos los modelos no paramétricos que propondremos. Para probar lo anterior consideremos una transformación monótona $\tau = g(\theta)$. Entonces,

$$(1.9) \quad P(\theta) = P(g^{-1}(g(\theta))) = P(g^{-1}(\tau)) = P^*(\tau)$$

donde ahora $P^* = P \circ g^{-1}$ es la curva característica del ítem relativa a τ .

Los modelos establecidos no determinan el origen ni las unidades de θ , es decir, si se utiliza otro origen y se cambian las unidades de θ y si se corrigen los parámetros adecuadamente, se obtendrán las mismas probabilidades $P(\theta)$. Si para simplificar nos restringimos a transformaciones lineales es claro que surge como pregunta natural cuáles son esas transformaciones. El teorema que sigue da respuesta a esa interrogante.

Teorema 1.3.1. a) Dado el ítem i , que se describe mediante el modelo de un parámetro, si se transforman θ y b_i sumando una constante a cada uno, la probabilidad permanece incambiada y el modelo se mantiene.

b) Dado el ítem i , que se describe mediante el modelo de dos parámetros, si se transforman θ , b_i y a_i mediante: $\theta' = m\theta + k$, $b'_i = mb_i + k$, $a'_i = \frac{a_i}{m}$ con $m > 0$ y k constantes arbitrarias, entonces la probabilidad permanece incambiada y el modelo se mantiene.

c) Dado el ítem i , que se describe mediante el modelo de tres parámetros, si se transforman θ , b_i y a_i mediante: $\theta' = m\theta + k$, $b'_i = mb_i + k$, $a'_i = \frac{a_i}{m}$ y c_i no se transforma, o sea $c'_i = c_i$ con

$m > 0$ y k constantes arbitrarias, entonces la probabilidad permanece incambiada y el modelo se mantiene.

1.3.1. Modelo de Rasch

El modelo logístico de un parámetro toma su nombre de su autor (Rasch, 1960). El modelo de Rasch es un caso particular del modelo de Birnbaum (1968) de tres parámetros, si consideramos que el parámetro de pseudo azar es 0 y la discriminación es igual para todos los ítems (en general se toma igual a 1). Si bien lo anterior es cierto, en su génesis, los modelos parten de supuestos diferentes. Este modelo tiene ventajas importantes pues la función de verosimilitud condicional toma una forma muy sencilla. Una característica importante de este tipo de modelo es que la puntuación total del sujeto es un estadístico suficiente para θ . Esto significa que la probabilidad de un vector de respuestas dado depende únicamente de la puntuación total.

Rasch (1960) en su libro *Probabilistic models for some intelligence and attainment tests* se acerca al análisis de los tests desde un marco probabilístico. Presenta tres modelos diferentes: uno para el número de errores en la lectura de un texto, otro para la velocidad de lectura y otro para los ítems de un test. Este último modelo será luego conocido como modelo de Rasch. Su presentación matemática se basa en una matriz de datos bidimensional obtenida al administrar n ítems a N examinados. Las respuestas son puntuadas en forma dicotómica $u_{ij} = 0, 1$, donde i indica el ítem y j al sujeto. A cada individuo se asocia un vector columna y notaremos por $U = [u_{ij}]$ la matriz $n \times N$, la matriz de datos antes mencionada. El vector que tiene los marginales de las columnas $u_{.j}$ es el vector de las puntuaciones totales r_j de los examinados y el que tiene los marginales de las filas $u_{.i}$ es el vector de las puntuaciones s_i de los ítems.

En este contexto Rasch emplea dos símbolos, $\eta_j \in [0, \infty]$, llamado la habilidad del examinado y $\delta_i \in [0, \infty]$, denominado la dificultad del ítem. Define la razón $\xi_{ij} = \frac{\eta_j}{\delta_i}$ y la llama parámetro situacional; se nota que la anterior razón también se puede definir como $\xi_{ij} = \eta_j \epsilon_i$ donde ϵ_i es el parámetro de facilidad. Es decir, $\delta_i = \frac{1}{\epsilon_i}$ sirve como parámetro de localización. Es bueno observar que en este caso la dificultad del ítem no es la misma dificultad que hemos visto en los modelos de dos y tres parámetros.

Rasch utilizó la función $f(x) = \frac{x}{1+x}$, que es el clásico ejemplo de biyección entre \mathbb{R} y $[0, 1]$, para poder conectar a la variable ξ_{ij} con una probabilidad. Se tiene entonces que,

$$(1.10) \quad P(u_{ij} = 1 | \xi_{ij}) = \frac{\xi_{ij}}{1 + \xi_{ij}} \quad \text{y} \quad P(u_{ij} = 0 | \xi_{ij}) = 1 - \frac{\xi_{ij}}{1 + \xi_{ij}} = \frac{1}{1 + \xi_{ij}}$$

luego, utilizando que $\xi_{ij} = \frac{\eta_j}{\delta_i}$ y sustituyendo obtenemos:

$$(1.11) \quad P(u_{ij} = 1 | \eta_j, \delta_i) = \frac{\frac{\eta_j}{\delta_i}}{1 + \frac{\eta_j}{\delta_i}} = \frac{\eta_j}{\delta_i + \eta_j} \quad \text{y} \quad P(u_{ij} = 0 | \eta_j, \delta_i) = \frac{1}{1 + \frac{\eta_j}{\delta_i}} = \frac{\delta_i}{\delta_i + \eta_j}$$

A partir de esto la forma general de la probabilidad de respuesta queda: $P(u_{ij} | \eta_j, \delta_i) = \frac{(\frac{\eta_j}{\delta_i})^{u_{ij}}}{1 + \frac{\eta_j}{\delta_i}}$

El parámetro situacional también puede definirse en términos de odds ratio, siendo:

$\xi_{ij} = \frac{P(u_{ij}=1 | \eta_j, \delta_i)}{P(u_{ij}=0 | \eta_j, \delta_i)} = \frac{\frac{\eta_j}{\delta_i + \eta_j}}{\frac{\delta_i}{\delta_i + \eta_j}} = \frac{\eta_j}{\delta_i}$. El parámetro situacional se mantiene igual ante cambios de escala, pues si multiplicamos la habilidad y la dificultad por una constante K se tiene que $\xi_{ij} = \frac{K\eta_j}{K\delta_i} = \frac{\eta_j}{\delta_i}$.

Al comienzo del apartado afirmamos que el modelo de Rasch es un caso particular del modelo de Birnbaum, esto es así pues si definimos $\theta_j = \ln \eta_j$ y $\beta_i = \ln \delta_i$, entonces,

$$(1.12) \quad P(u_{ij}|\theta_j, \beta_i) = \frac{(e^{\theta_j} e^{-\beta_i})^{u_{ij}}}{1 + e^{\theta_j} e^{-\beta_i}} = \frac{e^{(\theta_j - \beta_i)}}{1 + e^{(\theta_j - \beta_i)}}$$

El modelo de Rasch tiene dos propiedades que lo caracterizan. Una es que la puntuación total del sujeto es un estadístico suficiente para la puntuación del sujeto θ y la otra es que el modelo de Rasch posee la propiedad llamada de objetividad específica. Esta propiedad nos dice que si aplicamos n ítems que se ajustan al modelo de Rasch sobre subpoblaciones que tienen rasgos θ_1 y θ_2 , entonces para todo ítem i , si comparamos las razones de probabilidades en las subpoblaciones, esta será igual a la diferencia de los valores de los rasgos, es decir,

$$(1.13) \quad [\ln(\frac{P_i(\theta_1)}{Q_i(\theta_1)}) - \ln(\frac{P_i(\theta_2)}{Q_i(\theta_2)})] = \theta_1 - \theta_2$$

La propiedad es objetiva porque la relación no depende del ítem i utilizado y es específica pues depende de los n ítems que estiman el rasgo. Cualquier modelo de un parámetro para datos dicotómicos tiene esta propiedad. De los modelos logísticos, sólo el modelo de Rasch tiene las propiedades de objetividad específica y de poseer un estadístico suficiente para la aptitud. Esto permite que se mantengan las propiedades de la escala de intervalo para examinados e ítems. Un inconveniente es que los modelos de Rasch en ítems de respuesta múltiple, en general, no ajustan correctamente, aunque hay desarrollos que tratan de sobreponerse a este problema.

La importancia de estas propiedades radica en que se pueden usar para la estimación de los parámetros y la comparación de los individuos.

1.3.2. Función de información

Birnbaum (1968) define la función de información para una puntuación X como la razón entre el cuadrado de la pendiente de la regresión de $X|\theta$ y el error estándar de medida de X para un θ dado, es decir, $I(\theta, X) = \frac{(\frac{\partial \mu_{X|\theta}}{\partial \theta})^2}{\sigma_{X|\theta}^2}$. Cuando esa puntuación corresponde a la de un ítem i esta función se puede expresar como: $I_i(\theta) = \frac{[P'_i(\theta)]^2}{P_i(\theta)Q_i(\theta)}$.

Esto tiene conexión con la estimación por máxima verosimilitud de la habilidad y con la distribución muestral del estimador de θ . Cuando Birnbaum (1968) desarrolló su trabajo original, el método para obtener el estimador por máxima verosimilitud de la habilidad, no había sido desarrollado totalmente. Cramer demostró bajo condiciones generales que los estimadores máximo verosímiles (por ejemplo, $\hat{\theta}$), alcanzan la cota de Cramer-Rao dada por: $-E(\frac{\partial^2 \ln L}{\partial \theta^2})$ (donde L es la verosimilitud), son insesgados $E(\hat{\theta}) = \theta$ y tienen distribución asintótica normal con varianza el inverso de dicha cota, o sea, el inverso de la información de Fisher. Concretamente:

$$(1.14) \quad \hat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, -\frac{1}{E(\frac{\partial^2 \ln L}{\partial \theta^2})}) \quad \text{asintóticamente.}$$

Analicemos la información de Fisher en nuestro caso. Supongamos que un individuo responde a un test de n ítems y sus respuestas son puntuadas 0, si falla, o 1, si acierta. Además, supongamos conocidos los parámetros del ítem. Notemos por u_{ij} a la variable dicotómica que representa la respuesta del sujeto j al ítem i . Sea para el individuo j el vector de respuestas condicionado a su nivel de rasgo $U_j = (u_{1j}, u_{2j}, \dots, u_{nj}|\theta_j)$. Bajo el principio de independencia local las u_{ij} son independientes, a partir de esto la verosimilitud queda: $P(U_j|\theta_j) = \prod_{i=1}^n P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$ donde $P_{ij} = P_i(\theta_j)$.

Tomando logaritmo, $\ln L = \ln P(U_j|\theta_j) = \sum_{i=1}^n [u_{ij} \ln P_{ij} + (1 - u_{ij}) \ln Q_{ij}]$ en la última expresión omitimos θ_j en P y Q para simplificar la notación. Ahora,

$$(1.15) \quad \frac{\partial \ln L}{\partial \theta_j} = \sum_{i=1}^n [u_{ij} \frac{\partial \ln P_{ij}}{\partial \theta_j} + (1 - u_{ij}) \frac{\partial \ln Q_{ij}}{\partial \theta_j}] = \sum_{i=1}^n [u_{ij} \frac{P'_{ij}}{P_{ij}} + (1 - u_{ij}) \frac{Q'_{ij}}{Q_{ij}}]$$

$$(1.16) \quad \frac{\partial^2 \ln L}{\partial \theta_j^2} = \sum_{i=1}^n [u_{ij} \frac{P''_{ij} P_{ij} - P_{ij}^2}{P_{ij}^2} + (1 - u_{ij}) \frac{Q''_{ij} Q_{ij} - Q_{ij}^2}{Q_{ij}^2}] \quad \text{y tomando esperanzas}$$

$$E(\frac{\partial^2 \ln L}{\partial \theta_j^2}) = \sum_{i=1}^n [P_{ij} \frac{P''_{ij} P_{ij} - P_{ij}^2}{P_{ij}^2} + (1 - P_{ij}) \frac{Q''_{ij} Q_{ij} - Q_{ij}^2}{Q_{ij}^2}] = - \sum_{i=1}^n \frac{P_{ij}^2}{P_{ij} Q_{ij}} \quad \text{luego,}$$

$$(1.17) \quad I(\theta_j) = -E(\frac{\partial^2 \ln L}{\partial \theta_j^2}) = \sum_{i=1}^n \frac{P_{ij}^2(\theta_j)}{P_{ij}(\theta_j) Q_{ij}(\theta_j)}$$

que es la información del test y se descompone en la suma de las contribuciones de cada ítem, por lo que se define la información del ítem como: $I_i(\theta_j) = \frac{[P'_{ij}(\theta_j)]^2}{P_{ij}(\theta_j) Q_{ij}(\theta_j)}$. Entonces:

$$(1.18) \quad \hat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, \frac{1}{\sum_{i=1}^n \frac{P_{ij}^2(\theta_j)}{P_{ij}(\theta_j) Q_{ij}(\theta_j)}})$$

esto nos permite establecer, a un nivel dado θ_j , intervalos de confianza para θ_j para un nivel α mediante $\hat{\theta} \pm z_\alpha \frac{1}{\sqrt{I(\theta)}}$, donde z_α se determina a partir de la distribución normal.

Nos interesa ver cómo queda la función de información para cada uno de los modelos. En todos los casos consideramos $D=1$.

	Modelo normal	Modelo 1P	Modelo 2P	Modelo 3P
$P'_i(\theta)$	$a_i \frac{e^{\frac{a_i^2(\theta-b_i)^2}{2}}}{\sqrt{2\pi}}$	$P_i(\theta) Q_i(\theta)$	$a_i P_i(\theta) Q_i(\theta)$	$\frac{a_i [P_i(\theta) - c_i] Q_i(\theta)}{1 - c_i}$
$I(\theta)$	$\sum_{i=1}^n a_i^2 \frac{e^{\frac{a_i^2(\theta-b_i)^2}{2}}}{2\pi P_i(\theta) Q_i(\theta)}$	$\sum_{i=1}^n P_i(\theta) Q_i(\theta)$	$\sum_{i=1}^n a_i^2 P_i(\theta) Q_i(\theta)$	$\sum_{i=1}^n \frac{a_i^2 [P_i(\theta) - c_i]^2 Q_i(\theta)}{P_i(\theta) (1 - c_i)^2}$

Pasemos ahora a comentar algunos hechos interesantes de la función de información de los ítems. Ya vimos que la suma de las funciones de información de los ítems da la función de información del test. Esta propiedad nos permite, por lo tanto, si se tuvieran suficientes ítems, llegar a la función de información del test que deseamos, adicionando ítems convenientes al test. Actualmente, la función de información de los ítems es de las herramientas más utilizadas por quienes construyen tests, permitiéndoles mediante la combinación de los ítems obtener tests ajustados a sus necesidades. La función de información del ítem es un poderoso instrumento para el análisis de los ítems, no solamente por la cantidad de información que el ítem aporta a la medida de θ , sino también a qué nivel de θ aporta dicha información. También la función de información de los ítems se utiliza mucho en los tests adaptativos computarizados.

El valor de rasgo que aporta máxima información para un ítem dado es para el modelo de uno y dos parámetros $\theta = b_i$, con información en ese punto igual a 0.25 en el modelo de un parámetro, y $0.25a_i^2$ en el de dos. Para el modelo de tres parámetros $\theta = b_i + \frac{1}{a_i} \ln[\frac{1+\sqrt{1+8c_i}}{2}]$ y la información máxima es $\frac{a_i^2}{8(1-c_i)^2} [1 - 20c_i - 8c_i^2 + (1 + 8c_i)^{3/2}]$.

Una manera de medir la eficacia de dos tests para un valor dado de θ es mediante su eficiencia relativa definida como el cociente de sus funciones de información: $ER = \frac{I_1(\theta)}{I_2(\theta)}$

1.4. Modelos para ítems politómicos

El primero en introducir los modelos politómicos fue Samejima (1969, 1972) con su famoso modelo de respuesta graduada, el cual es una extensión del modelo logístico de dos parámetros. Se

le considera un modelo indirecto porque usa dos etapas para obtener las funciones de respuesta a cada categoría. Ha sido muy útil para la aplicación a test educacionales y es apropiado para respuestas politómicas ordenadas como, por ejemplo, escalas Likert o ítems de respuesta construida. En 1972, Bock presenta el modelo de respuesta nominal que es un modelo del tipo Thurstone, pues se deriva de las leyes de Thurstone (1927 y Thurstone y Chave 1929); Edwards y Thurstone (1952) a las que añade, fundamentalmente, un parámetro para describir al sujeto. Este modelo se basa en el supuesto de que en el sujeto se produce una reacción subjetiva ante cada estímulo y la respuesta observada depende de la relación entre el valor de dicha reacción y los parámetros del ítem.

El modelo de respuesta nominal o de categorías nominales propuesto por Bock (1972,1997) sirve para ítems con categorías de respuestas que no tienen un orden predeterminado. En su origen es una aproximación logística al método de las comparaciones pareadas de Thurstone (1927) pero también puede ser considerado como una extensión del modelo de elección de Luce (1959), o como un caso de regresión no lineal de variables latentes donde la función de enlace es la función logística multivariada. Un mayor detalle de esto se puede leer en Thissen y Steinberg (2010), que además proponen una nueva parametrización multidimensional.

Takane y de Leeuw (1987) proponen el modelo nominal multidimensional (con restricciones) como una versión de TRI del análisis factorial para variables nominales. Bolt y Johnson (2009) estudian respuestas extremas usando una versión del modelo multidimensional nominal.

Andrich (1978, 1988) presenta el modelo de escala de clasificación. También se desarrolla el modelo de crédito parcial y sus variantes (Masters y Wrigth, 1984; Andersen, 1973; Verhelst, Glas y de Vries, 1997). Muraki (1992) publica el modelo de crédito parcial generalizado que fue también propuesto por Yen al mismo tiempo y al que llamó modelo de crédito parcial de dos parámetros.

Adicionalmente, otros autores como McDonald y Embretson utilizan modelos multidimensionales para datos con respuestas politómicas. Zhang (2004b y 2007) extendió la teoría de las covarianzas condicionales al caso de ítems politómicos logrando fundamentos teóricos para el estudio de la dimensionalidad. Propone varios estimadores de la covarianza condicional, incluyendo el caso de diseños incompletos como los usados en NAEP. Abad, Olea y Ponsoda (2009) proponen un nuevo modelo para ítems de múltiple opción (*Restricted Samejima Multiple-Choice Model*) que es más explícito en los mecanismos de omisión de respuesta y lo usan para mejorar la estimación de los parámetros.

1.4.1. Modelo de categorías nominales

Formulación de Bock

Consideremos un ítem con m categorías nominales, la formulación original del modelo fue:

$$P(U = k|\theta, a, c) = P_k(\theta) = \frac{\exp z_k}{\sum_{i=0}^{m-1} \exp z_i}$$

donde $z_i = a_i\theta + c_i$ con $i = 0, 1, 2, \dots, m-1$. Este modelo usa la función logística multivariada y el proceso de respuesta de cada categoría (z_i) es una función lineal de la habilidad θ . Es decir,

$$P_k(\theta) = \frac{\exp(a_k\theta + c_k)}{\sum_{h=0}^{m-1} \exp(a_h\theta + c_h)}$$

Para que el modelo sea identificado deben agregarse dos restricciones. Bock(1972) sugiere

usar $\sum_{h=0}^{m-1} a_h = \sum_{h=0}^{m-1} c_h = 0$ que permiten reparametrizar el modelo a través de los parámetros α y γ (vectores de dimension m-1) definidos mediante las ecuaciones.

$$a = T\alpha \text{ y } c = T\gamma$$

donde T es una matriz de contraste del análisis de varianza definida por:

$$T_{m \times (m-1)} = \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \frac{1}{m} - 1 & \frac{1}{m} & \cdots & \frac{1}{m} \\ \frac{1}{m} & \frac{1}{m} - 1 & \cdots & \frac{1}{m} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} - 1 \end{pmatrix}$$

Los parámetros α y γ pueden ser estimados al tener dimensión m-1 cada uno. Pueden usarse otras matrices de contrastes para identificar el modelo; por ejemplo, Thiessen y Steinberg (1986) usan las restricciones $a_0 = c_0 = 0$. Es claro que este modelo no admite la no respuesta, ya que la suma de las probabilidades da 1 y que tampoco contempla la elección al azar de la respuesta. Thiessen y Cai (2016) dan una formulación general del modelo mediante:

$$P(U = k|\theta, a, S, c) = \frac{\exp z_k}{\sum_{i=0}^{m-1} \exp z_i}$$

donde $z_k = [a \circ S_{(k+1)}]' \theta + c_{(k+1)}$ y \circ representa el producto de Hadamard de matrices, θ es una la variable latente dimensión d , a es un vector de parámetros de discriminación (también de dimensión d), S es una matriz $d \times m$ de ponderadores y c es un vector de dimensión m de los interceptos.

Una propiedad interesante es que este modelo también sirve para ítems dicotómicos que tienen una CCI logística de dos parámetros. Para ello consideramos dos categorías, la respuesta incorrecta (0) y la respuesta correcta (1). Si aplicamos el modelo nominal tenemos:

$$P_0(\theta) = \frac{\exp(a_0\theta + c_0)}{\exp(a_1\theta + c_1) + \exp(a_0\theta + c_0)} \quad \text{y} \quad P_1(\theta) = \frac{\exp(a_1\theta + c_1)}{\exp(a_1\theta + c_1) + \exp(a_0\theta + c_0)}$$

Ahora, $P_1(\theta) = \frac{1}{1 + \exp((a_0\theta + c_0) - (a_1\theta + c_1))}$ y como $a_0 = -a_1$ y $c_0 = -c_1$ entonces,

$$P_1(\theta) = \frac{1}{1 + \exp(-2(a_1\theta + c_1))}$$

o sea, se obtiene un modelo logístico de 2P donde la discriminación es el doble de la pendiente de la opción correcta y la dificultad es $-c_1/a_1$.

Formulación de Thissen en 2010

Thissen y Steinberg (2010) parametrizan el modelo nominal de tal manera que facilita el análisis de ítems multidimensional con respuesta nominal. El modelo es:

$$P(U_i = k|\theta, \beta_{1i}, s_i, \beta_{0i}) = \frac{\exp z_k}{\sum_{h=0}^{m-1} \exp z_h}$$

donde los logits para cada categoría son: $z_h = [\beta_{1i} s_{i(k+1)}]' \theta + \beta_{0i(k+1)}$. Observemos que la diferencia con el modelo presentado antes es que, en este caso, s es un vector en vez de una matriz. Esto implica que todas las filas de la matriz S son iguales. Se imponen las restricciones $s_{11} = 0$, $s_{1(m-1)} = m - 1$ y $\beta_{0i1} = 0$ y se reparametriza mediante: $s_i = T\alpha_i$ y $\beta_{0i} = T\gamma_i$

Usando una matriz T adecuada se puede lograr un pasaje desde el modelo de respuesta nominal a un modelo de crédito parcial. Como matriz T se puede usar una base de Fourier

aumentada por una tendencia lineal. Así, T será una matriz de dimensión $m \times (m-1)$, que siendo $t_{ak} = \sin\left(\frac{\pi(k-1)(a-1)}{m-1}\right)$ toma la forma

$$T_F = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & t_{22} & \cdots & t_{2(m-1)} \\ 2 & t_{32} & \cdots & t_{3(m-1)} \\ \vdots & \vdots & \cdots & \vdots \\ m-1 & 0 & \cdots & 0 \end{pmatrix}$$

En esta matriz se impone que $\alpha_1 = 1$. Esta representación permite numerosas variantes del modelo nominal. Cuando se estiman $\beta_1, \alpha_1, \dots, \alpha_{(m-1)}$ y γ se obtiene el modelo nominal de rango completo. Si se restringe que $\alpha_1, \dots, \alpha_{(m-1)}$ sean iguales a 0 se obtiene una versión reparametrizada del modelo de crédito parcial generalizado. Esta matriz T permite crear modelos que están entre el nominal y el de crédito parcial generalizado. Si usamos una matriz T basada en la identidad podemos reparametrizar modelos que tienen restricciones de igualdad.

1.4.2. Modelo de escala de clasificación

Este modelo pertenece a la familia de modelos de Rasch (Rasch 1961, Andersen 1977, Andrich 1978, 2005) y puede aplicarse cuando la respuesta a los ítems son categorías ordenadas. Establece que la probabilidad de responder la categoría k ($k=0,1,\dots,m$) es:

$$P(U = k | \theta, b_i, \kappa_k, \varphi_k) = \frac{e^{\varphi_k(\theta - b_i) + \kappa_k}}{1 + \sum_{r=0}^m e^{\varphi_r(\theta - b_i) + \kappa_r}}$$

donde b_i es un parámetro de localización del ítem, φ_k y κ_k son llamadas funciones de puntaje y coeficientes de la categoría respectivamente. Andrich (1978) probó dos resultados para el caso unidimensional:

a) Si este modelo tiene estadísticos suficientes para la habilidad, entonces, $\varphi_{k+1} - \varphi_k = \varphi_k - \varphi_{k-1}$

b) Dos categorías adyacentes se pueden colapsar en una, solo si, $\varphi_k = \varphi_{k-1}$

Andersen (1977) no interpretó los parámetros del modelo, en tanto al siguiente año Andrich (1978) lo hizo en términos de umbrales y discriminaciones donde los umbrales se obtienen de particionar el continuo. Específicamente, si notamos por τ_r al umbral donde la probabilidad de responder a las categorías $r-1$ y r son iguales y notamos por c_r la discriminación en el umbral r entonces: $\kappa_0 = \kappa_m = 0$, $\kappa_k = -\sum_{i=1}^k c_i \tau_i$ y $\varphi_k = \sum_{i=1}^k c_i$

Se pueden hacer las siguientes observaciones:

a) Si las discriminaciones son positivas entonces las funciones de score son crecientes.

b) Si $c_k = 0$, entonces, $\varphi_k = \varphi_{k-1}$. Es decir, la condición que permite colapsar dos categorías adyacentes se puede interpretar como no tener discriminación entre ellas.

c) Si $c_i = 1$ para todo i ; $\kappa_k = -\sum_{i=1}^k \tau_i$ y $\varphi_k = k$, entonces, el modelo toma la forma:

$$P(U = k | \theta, b_i, \kappa_k, \varphi_k) = \frac{e^{k(\theta - b_i) - \sum_{i=1}^k \tau_i}}{1 + \sum_{r=0}^m e^{r(\theta - b_i) - \sum_{i=1}^r \tau_i}}$$

En este modelo es importante comprobar empíricamente que las categorías están ordenadas. En Andrich (2013) se puede ver una manera de comprobar dicho ordenamiento. Andersen (1973,1997) realiza la estimación condicional de los parámetros explotando el hecho que el puntaje total es un estadístico suficiente. Se han implementado otros métodos de estimación más generales. Un método, que mantiene la estimación condicional, tiene estimaciones consistentes, y permite datos faltantes, es la estimación condicional por parejas (Andrich y Luo, 2003) que se implementa en RUMM2030 (Andrich, Sheridan y Luo, 2014). Otros métodos utilizados incluyen

máxima verosimilitud conjunta, también conocida como estimación incondicional (Wright y Masters, 1982, Jansen et al., 1988, Wright, 1988) que se implementó en WINSTEPS (Linacre, 2002).

El método de máxima verosimilitud marginal (Bock y Aitkin, 1981), se implementa en ConQuest (Wu et al., 2007), y también se usa el método del valor esperado a posteriori (Bock y Mislevy, 1982).

1.4.3. Modelo de respuesta graduada

El primer modelo de respuesta graduada fue propuesto por Samejima (1969). Más tarde, propuso un marco más amplio para modelos de respuesta graduada, distinguiendo el caso homogéneo que pertenecen a modelos de ojiva normal y logísticos y el caso heterogéneo (Samejima, 1972).

Modelo general de respuesta graduada

Se supone que un proceso cognitivo, tal como resolver un problema, contiene un número m_i finito de pasos. Un ítem graduado asignará el puntaje u_i a los examinados que completan exitosamente el paso u_i , pero fallan al completar el paso $(u_i + 1)$. Sea $M_{u_i}(\theta)$ (Samejima, 1995) la función que da la probabilidad con la cual el examinado completa el paso u_i exitosamente, bajo las condiciones que el nivel de la habilidad del examinado es θ y el paso $(u_i - 1)$ ha sido completado exitosamente. Sea $(m_i + 1)$ el siguiente puntaje encima de m_i . Se establece que,

$$M_{u_i}(\theta) = \begin{cases} 1 & \text{para } u_i = 0 \\ 0 & \text{para } u_i = m_i + 1 \end{cases}$$

para todo θ . Para cada u_i se asume que $P_{u_i}(\theta)$ es no decreciente en θ (Samejima, 2016).

La función de respuesta a la categoría u_i es: $P_{u_i}(\theta) = [\Pi_{s \leq u_i} M_s(\theta)] [1 - M_{(u_i+1)}(\theta)]$. Esto provee el marco fundamental para el modelo general de respuesta graduada (Samejima, 2016). Sea $P_{u_i}^*(\theta) = \Pi_{s \leq u_i} M_s(\theta)$. Esta función es denominada de respuesta para puntajes categóricos acumulados (Samejima, 1995). La función de respuesta a puntajes categóricos también puede ser expresada como $P_{u_i}(\theta) = P_{u_i}^*(\theta) - P_{(u_i+1)}^*(\theta)$. El modelo general de respuesta graduada incluye varios modelos matemáticos específicos. Samejima (2016) comenta que el modelo debería ser examinado en los siguientes aspectos :

- que el principio detrás del modelo y el conjunto de supuestos estén de acuerdo con la realidad psicológica en cuestión.
- que se satisfaga la condición de un único máximo .
- que las modas de las funciones de respuesta a las distintas categorías estén ordenadas según los puntajes correspondientes.
- que se cumpla la aditividad de las funciones de respuesta.
- que el modelo pueda ser naturalmente generalizado a un modelo de respuesta continua.

La condición de que se tenga un único máximo se satisface si la función básica $A_{u_i}(\theta)$ definida por $A_{u_i}(\theta) = \frac{\partial}{\partial \theta} \log P_{u_i}(\theta) = \sum_{s \leq u_i} \frac{\partial}{\partial \theta} \log M_s(\theta) + \frac{\partial}{\partial \theta} \log [1 - M_{(u_i+1)}(\theta)]$ es estrictamente decreciente en θ , su asíntota superior es no negativa y su asíntota inferior es no positiva (Samejima 1969, 1972). Satisfacer esta condición asegura que la función de verosimilitud para cualquier

patrón de respuesta para esas categorías de respuestas tienen un único máximo sea este local o global. Usando la función básica, una condición suficiente, pero no necesaria, para el ordenamiento estricto de las modas de la función de respuestas categóricas es que $A_{(u_i-1)}(\theta) < A_{u_i}(\theta)$ para todo θ y $u_i = 1, 2, \dots, m_i$.

El caso homogéneo

El caso homogéneo del modelo de respuesta graduada representa a una familia de modelos con funciones $P_{u_i}^*(\theta)$, $u_i = 1, \dots, m_i$, que son idénticas en su forma y están posicionadas a lo largo del continuo de la habilidad en concordancia con el puntaje u_i del ítem. Entonces, es obvio que la aditividad de las funciones de respuesta categórica se cumple para los modelos del caso homogéneo. La función básica asintótica $\tilde{A}_{u_i}(\theta)$ ha sido definida en el caso homogéneo para $u_i = 1, 2, \dots, m_i$ por:

$$\tilde{A}_{u_i}(\theta) = \lim_{\lambda_{(u_i+1)} \rightarrow \lambda_{u_i}} A_{u_i}(\theta) = \frac{\partial}{\partial \theta} \log P_{u_i}(\theta) = \frac{\partial}{\partial \theta} \log \left[\frac{\partial}{\partial \theta} P_{u_i}^*(\theta) \right] - \frac{\frac{\partial^2}{\partial \theta^2} M_1(\theta - \lambda_{u_i})}{\frac{\partial}{\partial \theta} M_1(\theta - \lambda_{u_i})}$$

con λ_{u_i} tendiendo a cero para $u_i = 1$, creciente como función de u_i , e idéntica en forma para todo $u_i = 1, 2, \dots, m_i$ excepto en su posición en el continuo de θ (Samejima, 1969, 1972).

Samejima (1969, 1972) demostró que los modelos de ojiva normal y el modelo logístico pertenecen a esta clase de modelos y satisfacen la condición de un único máximo. En esos modelos, la función de respuesta categórica a los puntajes esta dada por:

$$P_{u_i}(\theta) = \frac{1}{[2\pi]^{1/2}} \int_{a_i(\theta-b_{u_i}+1)}^{a_i(\theta-b_{u_i})} \exp \left[\frac{-t^2}{2} \right] dt \quad y$$

$$P_{u_i}(\theta) = \frac{\exp [-Da_i(\theta - b_{u_i+1})] - \exp [-Da_i(\theta - b_{u_i})]}{[1 + \exp [-Da_i(\theta - b_{u_i})]] [1 + \exp [-Da_i(\theta - b_{u_i+1})]]} \quad (\text{Samejima, 2016})$$

El caso heterogéneo

Los modelos del caso heterogéneo de respuesta graduada se caracterizan por tener funciones de respuesta a las categorías $P_{u_i}^*(\theta)$, no todas idénticas en forma. Un ejemplo es el modelo de respuesta nominal de Bock (1972), ya que Samejima (1972) demostró que este modelo puede ser considerado como un modelo de respuesta graduada en el caso heterogéneo, si la respuesta nominal y los parámetros a_i del modelo de Bock están ordenadas, con una desigualdad estricta en, por lo menos, un par de ellos. Dos ejemplos son el modelo de crédito parcial y el modelo de crédito parcial generalizado.

Samejima (1995) propuso un modelo de aceleración que pertenece al caso heterogéneo que considera una situación donde el problema a resolver requiere un número de subprocesos para ser completados correctamente antes de alcanzar la solución. Se asume que hay más de un paso en el proceso y los pasos son observables. Puntajes graduados, o créditos parciales, de 1 a m_i son asignados al lograr exitosamente cada paso.

1.4.4. Modelo de crédito parcial y modelo de crédito parcial generalizado

El modelo de crédito parcial es una aplicación del modelo de Rasch para ítems dicotómicos. Si las respuestas al ítem i son alguna de las categorías $a = 0, 1, \dots, m_i$, entonces la probabilidad de que la respuesta de una persona al ítem i esté en la categoría x , en vez de en la categoría $x - 1$ es modelado por el modelo de Rasch para ítems dicotómicos (Masters, 2016).

Es el modelo más simple porque solamente incluye los parámetros de los ítems y las habilidades de las personas, lo cual lo hace fácil de aplicar. Ha sido utilizado rutinariamente en varios programas de evaluación. El modelo de crédito parcial difiere del modelo de respuesta graduada y del modelo de crédito parcial generalizado en que tanto este, como el modelo de escala de clasificación (el cual es un caso particular del modelo de crédito parcial), pertenecen a la familia de modelos de Rasch y por eso comparten características como la separabilidad de las personas y los parámetros de los ítems y la existencia de estadísticos suficientes. El modelo de crédito parcial (y sus casos particulares) es el único modelo en el cual la suma de los puntajes de los ítems es un estadístico suficiente para la habilidad de los sujetos. También es el único modelo que permite el ordenamiento estocástico de las personas por el puntaje total (Hemker et al., 1996; 1997; Ligtoet, 2012).

Cuando un ítem tiene más de dos categorías de respuestas, por ejemplo m categorías de respuestas ordenadas ($0 < 1 < 2 < \dots < m$), la probabilidad condicional de puntuar a en vez de $a - 1$ debe ser monótona creciente en el dominio de la habilidad. En el modelo de crédito parcial esto se hace usando el modelo de Rasch para respuesta dicotómica.

$$\frac{Pr\{U_{pi} = a\}}{Pr\{U_{pi} = a - 1\} + Pr\{U_{pi} = a\}} = \frac{e^{(\theta_p - b_{ia})}}{1 + e^{(\theta_p - b_{ia})}} = (a = 1, 2, \dots, m_i)$$

donde θ_p es la habilidad de la persona p y b_{ia} es un parámetro que gobierna la probabilidad de puntuar a en vez de $a - 1$ en el ítem i .

Condicionando un par de categorías adyacentes (se elimina toda otra respuesta posible de consideración), la anterior ecuación se enfoca en la comparación de las categorías $a - 1$ y a . Esta comparación local es el núcleo del modelo de crédito parcial y como demostró Masters (2010) es una característica esencial y definitoria de todos los modelos de la familia de Rasch.

Para aplicar el modelo de crédito parcial es conveniente reescribirlo como la probabilidad (incondicional) de cada posible resultado $0, 1, \dots, m_i$. Esto implica que al desempeño de la persona p en el ítem i le es asignado uno de los $m_i + 1$ puntajes disponibles, esto es, $\sum_{a=0}^{m_i} Pr\{U_{pi} = a\} = 1$.

La probabilidad de una persona p puntúe a en el ítem i se escribe como:

$$P\{U_{pi} = a\} = \frac{e^{\sum_{k=0}^a (\theta_p - b_{ik})}}{\sum_{h=0}^{m_i} e^{\sum_{k=0}^h (\theta_p - b_{ik})}} (a = 0, 1, \dots, m_i)$$

donde, $\sum_{k=0}^0 (\theta_p - b_{ik}) = 0$ y $\sum_{k=0}^0 (\theta_p - b_{ik}) = \sum_{k=1}^0 (\theta_p - b_{ik})$

El modelo de crédito parcial generalizado fue formulado por Muraki (1992) basándose en el modelo de crédito parcial de Masters. Este modelo relaja el supuesto de discriminaciones uniformes de los ítems del test. La diferencia entre estos dos tipos de modelos no es solamente la parametrización sino también el supuesto básico acerca de la variable latente.

El modelo de crédito parcial generalizado se escribe como: $P_{ia}(\theta) = \frac{e^{[\sum_{v=1}^a Z_{iv}(\theta)]}}{\sum_{c=1}^{m_i} e^{[\sum_{v=1}^c Z_{iv}(\theta)]}}$ y $Z_{ia}(\theta) = a_i(\theta - b_i + d_a)$ donde a_i es la pendiente, b_i es un parámetro de la ubicación del ítem y d_a es un parámetro de la categoría. El parámetro a_i indica el grado de cambio de la respuesta categórica al variar θ . Este concepto de la discriminación está relacionada con el índice de fiabilidad del ítem en la TCT (Muraki y Wang, 1992, Muraki y Muraki, 2016).

Como el número de categorías de respuestas es m , solamente $m - 1$ pueden ser identificados. Consecuentemente, algunos de los m parámetros pueden ser arbitrariamente fijados a algún valor fijo. Muraki y Muraki (2016) definen arbitrariamente $b_{i1} = 0$.

Los parámetros de la categoría del ítem $b_{ia} = b_i - d_a$ son puntos que pueden ocurrir en cualquier lugar de la escala de θ . Entonces, bajo el supuesto que $a_i > 0$, si $\theta = b_{ia}$, se cumple

, $P_{ia}(\theta) = P_{i,a-1}(\theta)$, si $\theta > b_{ia}$, entonces, $P_{ia}(\theta) > P_{i,a-1}(\theta)$ y si $\theta < b_{ia}$, entonces, $P_{ia}(\theta) < P_{i,a-1}(\theta)$.

Hay una indeterminación entre los b_i y los d_a en cada bloque de ítems que comparten el mismo conjunto de parámetros categóricos. Para obtener un conjunto de parámetros únicos, deben ser impuestas restricciones, por ejemplo, la media de los parámetros categóricos deben ser igual a una constante sobre los bloques. Las siguientes restricciones de ubicación pueden ser impuestas a través de bloques: $\sum_{a=2}^{m_i} d_a = 0$.

La suma de $Z_{iv}(\theta)$ para el ítem i en el modelo de crédito parcial generalizado puede ser escrita como: $Z_{ia}^+(\theta) = \sum_{v=1}^a Z_{iv}(\theta) = a_i [T_a(\theta - b_i) + K_a]$ donde $K_a = \sum_{v=1}^a d_v$. El valor esperado de la función de puntaje está dada por $\bar{T}_i(\theta) = \sum_{a=1}^{m_i} T_a P_{ia}(\theta)$.

$\bar{T}_i(\theta)$ es llamada función de respuesta del ítem y puede ser vista como la regresión de los puntajes del ítem sobre la escala de habilidad (Lord, 1980a; Muraki y Muraki, 2016).

1.5. Modelos de respuesta continua

Samejima (1973) extendió su modelo de respuesta graduada a respuestas a un intervalo abierto. Müller (1987) extendió el modelo de escala de clasificación de Andrich a respuestas a un intervalo cerrado, es decir, permite elegir los extremos. Tanto Samejima como Muller particionaron los intervalos dividiéndolos en mitades sucesivas y aplicando el modelo de respuesta graduada. Obtienen los modelos de respuesta continua como el caso límite cuando el número de intervalos tiende a infinito. Mellenbergh (2016) describe cómo los modelos de variable latente para el análisis de variables continuas han tenido otros desarrollos. Por ejemplo, modelos de análisis factorial fueron desarrollados para el análisis de correlaciones y covarianzas entre variables continuas (Bartholomew, 1995). La aplicación de estos modelos a ítems de respuestas continua produce los modelos de variable latente para ítems continuos (McDonald, 1982). Más adelante, los modelos de clase latente fueron desarrollados para el análisis de variables continuas (Gibson, 1959; Lazarsfeld y Henry, 1968).

Veamos el modelo, asumamos que un sujeto p contesta a un ítem continuo y que su respuesta depende de d variables latentes $\Theta_1, \Theta_2, \dots, \Theta_d$. También se asume que las respuestas del sujeto p al ítem i -ésimo están normalmente distribuidos con valores esperados τ_{pi} y varianza σ_{pi}^2 y que las respuestas condicionadas, al nivel de rasgo ($\Theta_1 = \Theta_{p1}, \dots, \Theta_d = \Theta_{pd}$) están distribuidas independientemente. Finalmente, se asume que la regresión de puntajes del ítem sobre las variables latentes pueden ser modelada como una función lineal de estas:

$$E(U_{pi} | \Theta_1 = \theta_{p1}, \Theta_2 = \theta_{p2}, \dots, \Theta_D = \theta_{pD}) = \tau_{pi} = b_i + a_{i1}\theta_{p1} + a_{i2}\theta_{p2} + \dots + a_{iD}\theta_{pD}$$

donde U_{pi} es el puntaje aleatorio del i -ésimo ítem del sujeto p ; b_i es el término independiente (intercepto); $a_{i1}, a_{i2}, \dots, a_{iD}$ son los coeficientes (pendientes) de la función de regresión del ítem i -ésimo; y $\theta_{p1}, \theta_{p2}, \dots, \theta_{pD}$ son los valores de las variables latentes del sujeto p .

El modelo factorial se obtiene de la ecuación anterior a partir de rasgos latentes continuos y asumiendo que la varianza de los puntajes del ítem es homogéneo con respecto a todos los sujetos. El modelo de clase latente es obtenido también de la ecuación anterior considerando a las variables latentes como variables latente dummy, que indican la pertenencia de los sujetos a determinada clase latente y asumiendo que la varianza al puntaje del ítem es homogéneo dentro de cada clase latente. El modelo de un factor de Spearman también es un caso particular tomando $d = 1$,

$$E(U_{pi} | \Theta = \theta_p) = \tau_{pi} = b_i + a_i\theta_p \quad \text{con} \quad Var(U_{pi} | \Theta = \theta_p) = \sigma_i^2$$

Este modelo, además del modelo factorial para variables continuas, es el modelo congenérico de Jöreskog (1971). Los parámetros b son comparables a los parámetros de dificultad de los modelos de TRI. Los parámetros a son comparables a los parámetros de discriminación de los modelos de TRI porque son pendientes de las funciones de regresión de los puntajes de los ítems en el rasgo latente.

Los casos especiales del modelo congenérico son obtenidos poniendo restricciones a los parámetros del ítem. Si $(a_i = a, b_i = b, \sigma_i^2 = \sigma^2, i = 1, 2, \dots, n)$ obtenemos la versión del modelo para ítems paralelos, si $(a_i = a, b_i = b, i = 1, 2, \dots, n)$ obtenemos el modelo esencialmente de ítems tau-equivalente. También, si $(a_i = a, \sigma_i^2 = \sigma^2, i = 1, 2, \dots, n)$ tenemos un modelo para puntajes de ítems continuos que es la contraparte del modelo de Rasch para respuestas a ítems dicotómicos (Mulaik, 1972). El modelo congenérico es comparable al modelo logístico de dos parámetros de Birnbaum para respuestas dicotómicas. Bajo el modelo de Birnbaum, la transformación logit de la probabilidad de dar una respuesta correcta es: $\ln \frac{\pi_{pi}}{1 - \pi_{pi}} = a_i(b_i - \theta_p) = b_i^* + a_i^* \theta_p$, donde $b_i^* = a_i b_i$, y $a_i^* = -a_i$.

Vemos que el modelo logístico de 2P tiene una estructura similar al modelo congenérico. Ambas ecuaciones son funciones lineales de una variable latente con la misma parametrización. El parámetro θ del modelo congenérico no es identificado. Una manera de hacerlo es fijar la media y la varianza del rasgo latente en la población igual a $E(\theta) = 0$ y $Var(\theta) = 1$.

Una diferencia entre el modelo logístico para respuestas dicotómicas y el modelo factorial de ítems continuos, es que el primero tiene solamente los parámetros a y b , y el último tiene parámetros a , b y σ^2 (Mellenbergh, 2016).

La generalización del modelo congenérico es el modelo factorial múltiple; la generalización del modelo logístico de 2P es el modelo de respuesta a ítems multidimensionales (Reckase, 2009). Aplicando la transformación logit a la probabilidad de dar una respuesta correcta al ítem bajo el modelo multidimensional para respuestas dicotómicas obtenemos la ecuación del modelo de factor múltiple para puntajes de ítems continuos.

Observemos que el modelo congenérico se apoya en supuestos muy fuertes como los mencionados al principio, pero además, a los efectos de obtener estimadores máximo verosímiles debe cumplirse que los puntajes tengan distribución normal multivariada lo que implica que también las variables latentes de los sujetos sean normalmente distribuidos. Estas suposiciones son muchas veces violadas, por eso se han desarrollado modelos que permiten estas violaciones.

Hessen y Dolan (2009) presentaron un modelo de un factor heterocedástico, el cual puede estimarse por máxima verosimilitud marginal. Molenaar, Dolan y Verhelst (2010) extienden el modelo a uno con distribución no normal de los parámetros de los sujetos y donde la varianza y el parámetro a pueden variar.

1.6. Estimación de modelos dicotómicos

Dedicaremos esta sección a presentar diferentes métodos de estimación para los modelos vistos hasta ahora. Decidimos presentarlos porque incluyen la descripción de las técnicas más utilizadas a lo largo de la TRI y como dijimos antes nos permite comparar estos algoritmos con los propuestos por nosotros.

Una vez definido el modelo a utilizar deben estimarse sus parámetros lo que usualmente se denomina calibración del ítem. Otro problema es estimar los rasgos o habilidades de los sujetos. Hay varios procedimientos para llevar a cabo la estimación de los ítems, los procedimientos basados en máxima verosimilitud, métodos bayesianos o los métodos Monte Carlo de cadenas

de Markov. También se pueden estimar los parámetros de los ítems suponiendo conocida la habilidad, o estimar la habilidad suponiendo conocidos los parámetros de los ítems. Asimismo, se pueden estimar al mismo tiempo los parámetros de los ítems y el rasgo.

La importancia de encontrar procedimientos que provean de estimadores que tengan buenas propiedades estadísticas es clara; pero además que sean eficientes computacionalmente, ya que si estimamos los parámetros de n ítems y el rasgo de N sujetos, si utilizamos un modelo de tres parámetros, habría $3n+N$ parámetros a estimar (aunque en realidad luego de fijar la escala se tendrán que estimar $3n+N-2$ parámetros) lo cual para valores grandes de N y n puede ser muy costoso desde el punto de vista computacional. A su vez, para tener una buena estimación se necesita que el tamaño de la muestra sea grande, que acentúa aún el problema anterior. Otro elemento que veremos en la estimación es la llamada pérdida de identificabilidad. Este problema ocurre cuando diferentes combinaciones de valores de los parámetros dan la misma distribución de los datos observados. Una revisión profunda de la identificabilidad se puede leer en San Martín (2016). Hay también resultados recientes que dan condiciones para chequear cuando las restricciones que se agregan son suficientes para la identificabilidad de los modelos. Es común en los modelos de efectos fijos poner como restricciones que la media de las habilidades sea 0 y la varianza 1.

Estimación por máxima verosimilitud

Muchos de los métodos de estimación utilizan la estimación por máxima verosimilitud. El método se basa en suponer que tenemos una función de densidad $p(x|\Theta)$ determinada por un conjunto de parámetros Θ . Tenemos una muestra aleatoria simple de tamaño n de observaciones con esta distribución, sea $X = \{x_1, \dots, x_n\}$. Como son independientes la densidad conjunta de esta muestra es: $p(X|\Theta) = \prod_{i=1}^n p(x_i|\Theta) = L(\Theta|X)$.

La función L se llama verosimilitud de los parámetros dado los datos. El método de máxima verosimilitud propone estimar los parámetros maximizando la función de verosimilitud, o sea los estimadores son: $\Theta^* = \operatorname{argmax}_{\Theta} L(\Theta|X)$.

En general, en vez de maximizar L se maximiza $\ln L$ pues es más sencillo. Los estimadores máximo verosímiles son los más usados en estadística debido a sus buenas propiedades formales. Estos estimadores son insesgados o asintóticamente insesgados y consistentes; son estadísticos suficientes o función de ellos; eficientes; asintóticamente normales y alcanzan la cota de Cramer-Rao. Concretamente su distribución en muestras grandes se puede enunciar como: los estimadores máximo-verosímiles $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ de los parámetros de una distribución dada por $f(x, \theta_1, \theta_2, \dots, \theta_k)$ para muestras de tamaño n , tienen, para muestras grandes, una distribución que se aproxima a la normal multivariante de medias $\theta_1, \theta_2, \dots, \theta_k$ y matriz R en la forma cuadrática en donde $r_{ij} = -E[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x, \theta_1, \theta_2, \dots, \theta_k)]$. Las varianzas y covarianzas de los estimadores son $(1/k)R^{-1}$. Esta propiedad se relaciona con la información de Fisher pues asintóticamente, $\hat{\theta} \sim N(\theta, \frac{1}{I(\theta)}) = N(\theta, -\frac{1}{E(\frac{\partial^2 \ln L}{\partial \theta^2})})$.

En el caso de la TRI maximizar la verosimilitud implica calcular los puntos estacionarios del gradiente. En la mayoría de los casos no se tiene una forma explícita de la solución por lo que deben emplearse métodos numéricos para hallarlos. El algoritmo de Newton-Raphson suele usarse combinado con el método de Fisher. En Luzardo y Forteza (2014) se puede encontrar el algoritmo de Newton-Raphson detallado para el caso de raíces de un gradiente basados en la matriz hessiana

Algoritmo EM

Otro algoritmo muy usado en la estimación es un algoritmo iterativo que fue propuesto por primera vez por Dempster, Laird y Rubin (1977) para obtener estimadores de máxima verosimilitud cuando las observaciones pueden ser vistas como datos incompletos o hay valores perdidos. Hay dos aplicaciones principales de este algoritmo: una es cuando hay valores *missing* debido a problemas o limitaciones en el proceso de observación, la segunda se da cuando optimizar la verosimilitud es analíticamente intratable, pero puede ser simplificada asumiendo la existencia y valores de parámetros no observados u ocultos.

El algoritmo se denomina EM porque se compone de un paso donde se toman esperanzas seguido por un paso de maximización. Supongamos que los datos observados X son generados por alguna distribución, llamaremos X a los datos incompletos. Supongamos que los datos completos vienen dados por $Z=(X,Y)$ y tienen densidad conjunta $p(z|\Theta) = p(x,y|\Theta) = p(y|x, \Theta)p(x|\Theta)$. Podemos definir una nueva función de verosimilitud llamada la verosimilitud para los datos completos, definida por: $L(\Theta|Z) = L(\Theta|X, Y) = p(X, Y|\Theta)$. Esta función es una variable aleatoria pues Y es desconocido y aleatorio. Podemos pensar entonces $L(\Theta|X, Y) = h_{X,\Theta}(Y)$, donde X y Θ en esta función son constantes, e Y una variable aleatoria.

En un primer paso, el algoritmo EM encuentra el valor esperado del logaritmo de la verosimilitud para los datos completos con respecto a los datos desconocidos Y , dado los valores observados X y los estimadores de los parámetros en ese momento. Esto es, se define: $Q(\Theta, \Theta^{(i-1)}) = E(\ln p(X, Y|\Theta)|X, \Theta^{(i-1)})$ donde $\Theta^{(i-1)}$ son los parámetros actuales usados para evaluar las esperanzas y Θ son los nuevos parámetros.

En la expresión previa X y $\Theta^{(i-1)}$ son constantes, Y es una variable aleatoria que tiene distribución $f(y|X, \Theta^{(i-1)})$ y Θ los parámetros a estimar.

Calculando la esperanza condicional tenemos:

$$E(\ln p(X, Y|\Theta)|X, \Theta^{(i-1)}) = \int_{y \in \Omega} \ln p(X, y|\Theta) f(y|X, \Theta^{(i-1)}) dy$$

donde $f(y|X, \Theta^{(i-1)})$ es la distribución marginal de los datos no observados y es dependiente de X y de los parámetros observados, Ω es el espacio de valores de y . El cálculo de esta esperanza se denomina paso E.

Luego, el segundo paso consiste en maximizar $Q(\Theta, \Theta^{(i-1)})$, o sea: $\Theta^i = \operatorname{argmax}_{\Theta} Q(\Theta, \Theta^{(i-1)})$

Una modificación del algoritmo EM consiste, en vez de maximizar Q , en encontrar algún Θ^i tal que $Q(\Theta^i, \Theta^{(i-1)}) > Q(\Theta, \Theta^{(i-1)})$. Esta modificación da origen al algoritmo EM generalizado (GEM). Dempster et al. (1977) prueban que en cada iteración crece el logaritmo de la verosimilitud y que el algoritmo converge a un máximo local de la verosimilitud. También dan tasas y velocidades de convergencia.

Revisión de los procedimientos

En esta sección se pretende hacer un recuento de los métodos y circunstancias en que se aplican. Si se estiman los parámetros de los ítems suponiendo conocida la habilidad se puede emplear estimación por máxima verosimilitud, estimación bayesiana y estimación por mínimo χ^2 . Si se estima la habilidad suponiendo conocidos los parámetros de los ítems, hay métodos basados en máxima verosimilitud y métodos bayesianos.

En múltiples casos se estiman los parámetros de los ítems y la habilidad conjuntamente. En este caso destacan la estimación conjunta por máxima verosimilitud, por máxima verosimilitud condicional y por máxima verosimilitud marginal. También está la estimación bayesiana.

Además, existen métodos alternativos que pretenden dar procedimientos de bajo coste computacional. Dentro de estos está el que utiliza la correlación biserial puntual y la dificultad del ítem (este método se verá luego como parte de los algoritmos para encontrar estimaciones iniciales).

También está el método de Urry para estimar los parámetros del modelo 3P. Verhelst y Molenaar (1988) usan regresión logística con métodos iterativos.

En la estimación aparecen ciertos detalles a tener en cuenta. El primero, está relacionado con el problema de la identificación, que ya revisamos. Para resolver este problema se suele fijar la media de θ a cero y su varianza a 1 (por consiguiente también se transformará la dificultad del ítem).

Se han probado resultados sobre la convergencia de los estimadores:

a) Si los parámetros de los ítems son conocidos, el estimador máximo verosímil $\hat{\theta}$ de la habilidad converge a θ si el número de ítems es suficientemente grande.

b) Si la aptitud es conocida, los estimadores máximo verosímiles de los parámetros de los ítems convergen al valor de esos parámetros si el número de examinados es suficientemente grande.

c) Si se estiman conjuntamente los parámetros de los ítems y la habilidad, los parámetros de los ítems son estructurales y los de la habilidad incidentales (pues crece el número de parámetros al aumentar la muestra). En este caso los estimadores de los parámetros de los ítems no convergen a sus parámetros cuando la muestra de examinados se incrementa. Sin embargo, se logra consistencia si además se hace que el número de ítems tienda a infinito. Es decir, para la estimación simultánea se necesitan grandes muestras de examinados y de ítems.

Lo que nos interesa resaltar aquí es la complejidad de los algoritmos de estimación. Para ello veremos algunos detalles. En el caso a), donde la habilidad se considera conocida, se aplica el método de máxima verosimilitud tanto a los modelos de ojiva normal, como logísticos de 1, 2 y 3 parámetros. No hay una solución explícita para determinar los puntos estacionarios de la verosimilitud por lo que se utiliza el método de Newton Raphson. Esto da origen a un algoritmo iterativo en el cual además suele emplearse el método de Fisher, pues en general, converge más rápidamente que Newton Raphson. Kale (1962) mostró que los valores iniciales deben estar en un entorno de los parámetros y deben estar basados en estimadores consistentes para que el proceso iterativo converja.

Para ejemplificar si queremos estimar el modelo 2P suponiendo las habilidades conocidas, en vez de estimar directamente a y b se estiman $\zeta = -ab$ y $\lambda = a$ ya que linealizan el exponente del modelo. Se consideran k grupos que tienen cada uno rasgo θ_j y que están constituidos por f_j sujetos cada uno. El algoritmo puede resumirse en los siguientes pasos:

1- Se comienza con estimaciones primarias de $(\hat{\zeta}_1, \hat{\lambda}_1)$.

Hay varios esquemas para elegir los valores iniciales, por ejemplo en el caso del modelo de ojiva normal podemos usar las ecuaciones de Tucker (1946). Bajo el modelo de dos parámetros los valores de tendencia e intercepto 1 y 0 muchas veces funcionan adecuadamente.

2- Si en el tiempo t tenemos $(\hat{\zeta}_t, \hat{\lambda}_t)$, a partir de estos calculamos $\hat{\zeta}_t + \hat{\lambda}_t \theta_j$ para todos los niveles de habilidad.

3- Se calcula \hat{P}_j , \hat{Q}_j , y $w_j = \hat{P}_j \hat{Q}_j$.

4- Se calcula $(\hat{\zeta}_{t+1}, \hat{\lambda}_{t+1})$ mediante:

$$(1.19) \quad \begin{pmatrix} \hat{\zeta} \\ \hat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \hat{\zeta} \\ \hat{\lambda} \end{pmatrix}_t + \begin{pmatrix} \sum_{j=1}^k f_j w_j & \sum_{j=1}^k f_j w_j \theta_j \\ \sum_{j=1}^k f_j w_j \theta_j & \sum_{j=1}^k f_j w_j \theta_j^2 \end{pmatrix}_t^{-1} \begin{pmatrix} \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)}{\hat{P}_j \hat{Q}_j} \\ \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)}{\hat{P}_j \hat{Q}_j} \theta_j \end{pmatrix}_t$$

5- Este proceso se continúa hasta que $|\Delta \hat{\zeta}_t|$ y $|\Delta \hat{\lambda}_t|$ son arbitrariamente chicos.

6- A partir de las estimaciones de $\hat{\zeta}$ y $\hat{\lambda}$ obtenemos $\hat{a} = \hat{\lambda}$ y $\hat{b} = -\frac{\hat{\zeta}}{\hat{\lambda}}$

Berkson (1955) halló las varianzas asintóticas de los estimadores:

$$s_{\bar{\lambda}}^2 = \frac{1}{\sum_{j=1}^k f_j w_j (\bar{\theta} - \theta_j)^2} = s_a^2 \quad \text{donde} \quad \bar{\theta} = \frac{\sum_{j=1}^k f_j w_j \theta_j}{\sum_{j=1}^k f_j w_j}$$

$$s_{\hat{\zeta}}^2 = \frac{1}{\sum_{j=1}^k f_j w_j} + \bar{\theta}^2 s_{\bar{\lambda}}^2 \quad \text{y} \quad s_b^2 = \frac{1}{a^2} \left[\frac{1}{\sum_{j=1}^k f_j w_j} + s_a^2 (\hat{b} - \bar{\theta})^2 \right]$$

Un test de bondad de ajuste obtenido por Garwood (1941) es $\chi^2 = \sum_{j=1}^k f_j w_j \frac{(p_j - P_j)^2}{P_j Q_j}$ que tiene distribución asintótica chi cuadrado con k-2 grados de libertad.

Este esquema funciona correctamente para los modelos de uno y dos parámetros, sin embargo para el modelo de 3 parámetros es muy inestable debido a las propiedades de la superficie de máxima verosimilitud. Este es un tema de trabajo aun en día de los matemáticos que trabajan en optimización.

Otro procedimiento aunque no muy usual es el de estimación por mínimo χ^2 . Este procedimiento fue propuesto por Berkson (1955), donde se maximiza el ajustes entre los valores esperados y observados minimizando el estadístico χ^2 .

Los procedimientos de estimación de la habilidad cuando se conocen los parámetros de los ítems son bastante más sencillos. En este caso el proceso iterativo toma la forma

$$[\hat{\theta}_j]_{t+1} = [\hat{\theta}_j]_t - \left[\frac{\partial^2 L}{\partial \theta_j^2} \right]_t^{-1} \left[\frac{\partial L}{\partial \theta_j} \right]_t$$

Por ejemplo en el modelo 3P el algoritmo queda:

$$[\hat{\theta}_j]_{t+1} = [\hat{\theta}_j]_t + \left[\sum_{i=1}^n -\lambda_i^2 Q_{ij} \left[\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}} \right] \right]_t^{-1} \left[\sum_{i=1}^n \lambda_i \frac{(u_{ij} - P_{ij})(P_{ij} - c_i)}{(1 - c_i) P_{ij}} \right]_t$$

y la varianza asintótica es:

$$s_{\hat{\theta}_j}^2 = \frac{1}{-E\left(\frac{\partial^2 \ln L}{\partial \theta_j^2}\right)} = \frac{1}{\sum_{i=1}^n \lambda_i^2 Q_{ij} \left[\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}} \right]} = \frac{1}{\sum_{i=1}^n \alpha_i^2 Q_{ij} \left[\frac{(P_{ij} - c_i)^2}{(1 - c_i)^2} \frac{1}{P_{ij}} \right]}$$

Cuando interesa la estimación conjunta de parámetros del ítem y habilidad hay varios procedimientos. Uno de ellos lo hace mediante máxima verosimilitud. El argumento central está dado por el paradigma de Birnbaum que simplifica los algoritmos a utilizar. Este tipo de estimación exige la eliminación de las puntuaciones extremas, tanto para ítems como para sujetos. Por lo tanto, se deben eliminar los sujetos que acierten o fallen todos los ítems. Tampoco puede usarse cuando los ítems pueden dividirse en dos clases, de tal forma que los sujetos acierten todos los ítems de la primera clase y fallen los de la segunda. Su principal limitación es que si tenemos N sujetos y n ítems las estimaciones son inconsistentes para n fijo cuando $N \rightarrow \infty$. Son consistentes si $N \rightarrow \infty$, $n \rightarrow \infty$ y $\frac{N}{n} \rightarrow \infty$. Como en la mayoría de los casos se tiene un número limitado de ítems, las estimaciones conjuntas de los ítems no son consistentes y tampoco asintóticamente insesgadas. Aplicando las derivadas de primer orden al logaritmo de la verosimilitud se obtiene un sistema de $2n+N$ derivadas igualadas a 0, luego se usa el algoritmo de Newton Raphson.

El algoritmo toma la forma $\psi_{t+1} = \psi_t - H_t^{-1}(\nabla L)_t$ donde ψ es el vector de dimensión $2n+N$ que contiene los parámetros a estimar; H la matriz hessiana (dimensión $2n+N \times 2n+N$) y ∇L el vector gradiente del logaritmo de la verosimilitud. La dimensionalidad del problema es grande así que se han tratado de encontrar formas de reducirla y con este objeto se hacen tres supuestos:

1- dado que los examinados se toman aleatoriamente de la población, cada examinado es un objeto independiente, luego las derivadas cruzadas entre pares de examinados tiene esperanza cero.

2- no hay motivo para que exista covariación entre un sujeto y los parámetros de los ítems, así que se asumen independientes y por lo tanto se eliminan las derivadas cruzadas.

3- se asume que los ítems son independientes y, por lo tanto, se eliminan las derivadas cruzadas entre ítems diferentes.

Después de estas simplificaciones la matriz hessiana sólo está compuesta de submatrices 2x2 para cada ítem y elementos en la diagonal para los sujetos.

Aquí entra en juego el paradigma de Birnbaum (1968) que propone un proceso de estimación en dos etapas. En la primera etapa se estiman los parámetros de los ítems asumiendo que las habilidades son conocidas. Para iniciar el proceso iterativo se toman estimaciones crudas de las habilidades, en general, es el puntaje estandarizado del sujeto. Como por los supuestos se pueden considerar los ítems separadamente, se resuelve los estimados iterativamente por:

$$(1.20) \quad \begin{pmatrix} \hat{\zeta} \\ \hat{\lambda} \end{pmatrix}_{t+1} = \begin{pmatrix} \hat{\zeta} \\ \hat{\lambda} \end{pmatrix}_t - \begin{pmatrix} E(\frac{\partial^2 L}{\partial \zeta_i^2}) & E(\frac{\partial^2 L}{\partial \zeta_i \partial \lambda_i}) \\ E(\frac{\partial^2 L}{\partial \lambda_i \partial \zeta_i}) & E(\frac{\partial^2 L}{\partial \lambda_i^2}) \end{pmatrix}_t^{-1} \begin{pmatrix} \frac{\partial L}{\partial \zeta_i} \\ \frac{\partial L}{\partial \lambda_i} \end{pmatrix}_t$$

A partir de resolver estas ecuaciones se tienen las estimaciones de los parámetros de los ítems y suponiéndolos conocidos podemos estimar las habilidades mediante $[\hat{\theta}_j]_{t+1} = [\hat{\theta}_j]_t - [E(\frac{\partial^2 L}{\partial \theta_j^2})]_t^{-1} [\frac{\partial L}{\partial \theta_j}]_t$

En este punto un primer ciclo del paradigma de Birnbaum se ha completado y tenemos estimados de los parámetros de los ítems ζ_i y λ_i para todo i y de las habilidades de los sujetos θ_j para todo j . Como la métrica de θ es única a menos de transformaciones monótonas, aunque no sería necesario fijar la escala, algunos programas como el LOGIST lo hacen mediante la media y la desviación típica de las habilidades. Se tiene entonces, el score estándar: $\theta_j^* = \frac{(\hat{\theta}_j - \bar{\theta})}{s_{\hat{\theta}}}$. También se ajustan los parámetros de los ítems mediante $b_i^* = \frac{(\hat{b}_i - \bar{\theta})}{s_{\hat{\theta}}}$ y $a_i^* = s_{\hat{\theta}} \hat{a}_i$.

Por último, se necesita un criterio de parada a los efectos de finalizar los ciclos. Este criterio es arbitrario. Por ejemplo, Wood, Wingersky y Lord (1976) proponen que la diferencia entre dos ciclos sucesivos debe ser menor a un valor específico.

Para la estimación del modelo de Rasch se desarrollaron dos procedimientos, un basado en máxima verosimilitud condicional y otro en máxima verosimilitud conjunta.

Rasch indica en 1960 que las expresiones de probabilidad condicional podrían ser usadas con propósitos de estimación. Rasch provee un algoritmo y su implementación, pero no funcionó bien en la práctica. Andersen dio un método basado en estimación por máxima verosimilitud basada en probabilidades condicionales. Este procedimiento solo se aplica para estimar los parámetros de los ítems. Una vez que los ítems son calibrados se estima la habilidad basándose en la disponibilidad de estadísticos suficientes para la habilidad. En este procedimiento se calcula la función de verosimilitud condicionada a la puntuación total. Además, se soluciona el problema que se daba en la estimación conjunta respecto a la cantidad de parámetros incidentales, pues la cantidad de posibles puntuaciones es igual a la cantidad de ítems más 1.

Como ya sabemos hay una indeterminación en la definición de la métrica. Andersen (1973 y 1977) impone la restricción $\sum_{i=1}^n \beta_i = 0$ a los efectos de determinar la escala. Wright y Douglas (1977) implementaron esta restricción sustrayendo la derivada de un ítem de las restantes, por ejemplo el ítem n . El procedimiento por máxima verosimilitud conjunta estima tanto los parámetros de los ítems como las habilidades. Este algoritmo tiene dos procedimientos iterativos que utilizan Newton Raphson uno para estimar las dificultades de los ítems y otro para estimar las habilidades.

Para los modelos que no son Rasch hay que recurrir a otros métodos uno de ellos es la estimación usando máxima verosimilitud marginal. Aunque este proceso no se considera una

técnica bayesiana el proceso de marginalización incorpora el teorema de Bayes en su matemática. Sea $f(\theta|\tau)$ la densidad de probabilidad de la habilidad en la población (distribución a priori) y τ el vector de parámetros de la densidad, se supondrá que la distribución a priori es la misma para todos los examinados. Sea U_j el vector de respuestas del sujeto j y ξ el vector de parámetros de los ítems. Se aplica el teorema de Bayes para calcular la distribución a posteriori de θ_j . Se tiene que:

$$(1.21) \quad P(\theta_j|U_j, \tau, \xi) = \frac{P(U_j|\theta_j, \xi)f(\theta|\tau)}{\int P(U_j|\theta_j, \xi)f(\theta_j|\tau)d\theta}$$

donde $P(U_j|\theta_j, \xi)$ es la verosimilitud. Por la independencia local, $P(U_j|\theta_j, \xi) = \prod_{i=1}^n P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$

La integral que aparece en el denominador es la probabilidad marginal (no condicionada a θ_j) del vector de respuestas con respecto a los parámetros de los ítems. Para resolver finalmente el problema destacan dos posibles algoritmos, el algoritmo de Bock y Lieberman y el algoritmo de Bock y Aitkin. En el algoritmo de Bock y Lieberman se maximiza la distribución marginal. En las derivadas respecto a los parámetros aparece una integral que surge de la marginalización. Para aproximar dichas integrales se aplica la cuadratura de Hermite-Gauss. Si se tiene una densidad continua con momento finito $f(\theta|\tau)$ se puede aproximar mediante un histograma. Luego la integral puede ser reemplazada por la suma de las áreas de un número finito de rectángulos. Supongamos que tenemos q intervalos y sea X_k el punto medio del intervalo k , al que llamaremos nodo. Cada nodo tiene un peso $A(X_k)$ que toma en cuenta la altura de la función de densidad en un entorno de X_k . Bock y Lieberman (1970) modifican los nodos y pesos de Stroud y Sechrest para el caso de la densidad normal típica. Sustituyendo la aproximación de la integral por su fórmula de cuadratura pueden aplicar el método de N-R para estimar los $3n$ parámetros del test simultáneamente. Como en el caso de la estimación conjunta, los u_{ij} son reemplazados por su esperanza, o sea se aplica el método de Fisher. El método no es atractivo desde el punto de vista computacional, ya que requiere invertir una matriz $3n \times 3n$ y es por esto que este procedimiento se limita a test con muy pocos ítems.

En el algoritmo de Bock y Aitkin los autores asumen que los ítems son independientes, los sujetos también y que son independientes los sujetos de los ítems, es decir, utilizan los mismos supuestos que Birnbaum en su paradigma para la estimación conjunta. Bock y Aitkin generan unos datos que dependen de probabilidades condicionales y a los que se les llama datos artificiales. Los parámetros de los ítems se estiman usando el método de Newton Raphson o Fisher. Los estimadores iniciales de los parámetros que suelen emplearse se basan en el coeficiente de correlación biserial entre θ y la variable de respuesta U y en la dificultad de ítem.

En este método los datos artificiales dependen de los estimadores de los parámetros de los ítems, por lo tanto, debe establecerse una metodología donde los datos artificiales y los parámetros de los ítems se pueden obtener iterativamente. Para esto Bock y Aitkin (1981) emplearon el algoritmo EM que como ya dijimos es un procedimiento iterativo que sirve para encontrar los estimadores máximo verosímiles en presencia de variables no observables y que consta de dos pasos: E (esperanza) y M (maximización). En nuestro caso queremos estimar los parámetros de los ítems en presencia de una variable aleatoria no observable (θ). Para realizar inferencia acerca de θ se usa alguna representación observable basada en la respuesta de los sujetos. Sea el par (U, θ) los datos no observados (datos completos) y el vector U los datos observados (datos incompletos). Sea $f(U, \theta|\xi)$ la densidad de probabilidad conjunta de (U, θ) y ξ^p los parámetros estimados en el ciclo p , entonces ξ^{p+1} se calculan maximizando $E(\ln f(U, \theta|\xi)|U, \xi^p)$ con respecto a ξ . Este proceso se repite hasta satisfacer el criterio de convergencia. Los dos pasos serían entonces:

Paso E: Calcular $E(\ln f(U, \theta|\xi)|U, \xi^p)$.

Paso M: Calcular ξ^{p+1} como aquel que maximiza la esperanza posterior.

Se encuentran además métodos bayesianos, donde por ejemplo si se desea estimar los parámetros de los ítems cuando los parámetros de habilidad son conocidos entonces se considera la distribución a priori de los parámetros del i -ésimo ítem, como $g(\xi, \eta)$, y se asume además que todos los ítems tienen la misma distribución a priori. La distribución a posteriori de la verosimilitud es: $g(\xi|U, \theta, \eta) \propto L(U|\xi, \theta)g(\xi, \eta)$ donde el signo de proporcionalidad toma el sentido usual de la estadística bayesiana y

$$L(U|\xi, \theta) = \prod_{j=1}^N \prod_{i=1}^n P_i(\theta_j)^{u_{ij}} Q_i(\theta_j)^{1-u_{ij}} = \prod_{j=1}^N P(U_j|\theta_j, \xi)$$

Esta distribución puede usarse para estimar los parámetros. Mislevy (1986b) presentó un procedimiento de estimación bayesiano que es una extensión de la solución marginal de Bock y Aitkin (1981). Emplea un procedimiento en dos etapas en el cual la información a priori es especificada en un modo jerárquico. Comienza con la distribución conjunta a priori de todos los parámetros: $g(\theta, \tau, \xi, \eta) = \prod_{j=1}^N g(\theta_j|\tau) \prod_{i=1}^n g(\xi_i|\eta)g(\tau)g(\eta)$

La distribución a posteriori aplicando el teorema de Bayes es:

$$g(\theta, \tau, \xi, \eta|U) \propto L(U|\theta, \xi)g(\theta|\tau)g(\tau)g(\xi|\eta)g(\eta)$$

Esta ecuación se marginaliza respecto a θ y Mislevy sugiere que también se haga respecto a η , pues los considera fuera de interés y los trata como parámetros de ruido. A partir de esto:

$$g(\xi, \tau|U) \propto \int \int L(U|\theta, \xi)g(\theta|\tau)g(\tau)g(\xi|\eta)g(\eta)d\theta d\eta \propto L(U|\xi, \tau)g(\xi)g(\tau).$$

donde $L(U|\xi, \tau)$ es la verosimilitud marginal. En la estimación marginal bayesiana se asume que los ítems son independientes lo que permite proceder ítem a ítem. También se puede proceder vía el algoritmo EM. Dentro de los métodos bayesianos para estimar las habilidades encontramos la estimación modal bayesiana y del valor esperado a posteriori.

1.7. Modelos multidimensionales

Los modelos multidimensionales de TRI tratan de modelizar la probabilidad de las puntuaciones de los ítems condicionadas a un rasgo perteneciente a un espacio latente multidimensional. Es interesante que los modelos multidimensionales de TRI son un caso particular del análisis factorial o de los modelos de ecuaciones estructurales usando variables no estandarizadas (Raykov y Marcoulides, 2011) y también del caso general de estos métodos cuando se trabaja con variables estandarizadas (McDonald, 1999).

Horst (1965) realizó los fundamentos que permitieron el desarrollo de los modelos de TRI multidimensional al sugerir realizar el análisis factorial con la matriz de datos en vez de usar la matriz de correlaciones. Para analizar ítems dicotómicos multidimensionales McDonald (1967) recomienda usar funciones no lineales en el análisis factorial.

McDonald (1982) clasifica los modelos de rasgo latente en:

1. estrictamente lineales, donde la densidad de las respuestas condicionada al rasgo es lineal, tanto en los coeficientes como en los rasgos latentes.
2. lineales en sentido amplio, donde la densidad condicional es lineal en los coeficientes, pero no en los rasgos latentes.
3. estrictamente no lineales, donde la densidad condicional no se puede expresar linealmente en sentido amplio con un número finito de términos.

El último caso es el del modelo de respuesta al ítem para respuestas dicotómicas donde la función de respuesta está definida por la función logística acumulada y la ojiva normal.

La relación entre los modelos de TRI multidimensional y el análisis factorial quedo establecida firmemente en los trabajos de Christofferson (1975), Muthen (1978), y Bock y Aitkin (1981).

El modelo de ojiva normal y el logístico explotan el hecho que pueden ser expresados como una suma de polinomios ortogonales (Courant y Hilbert, 1953). El aplicar el análisis armónico permitió a McDonald (1962a, 1967) desarrollar un método para su estimación. El procedimiento de estimación fue implementado en el programa NOHARM (normal-ogive harmonic analysis robust method), quien utiliza un polinomio de grado 3 para aproximar al modelo de ojiva normal. En Fraser y McDonald (1988) se dan detalles de esta implementación.

Considérense dos vectores de variables aleatorias ν y η de dimensiones $n \times 1$ y $d \times 1$, entonces:

$$P(\nu = u) = \int \dots \int_{R^d} g(\nu = u | \eta = h) f(h) dh$$

donde u y h son dos realizaciones de las variables ν y η , respectivamente, $f(h)$ es la función de densidad de las variables aleatorias η y la integración se realiza con respecto a η . Se denominan modelos de rasgos latentes multidimensionales a aquellos modelos en los que:

a) Las únicas variables que pueden ser observadas son las ν , mientras que las variables aleatorias η son inobservables y se asume que son continuas.

b) Las variables ν son independientes entre sí para un valor fijo de la variable η , digamos $\eta = h$. Es decir, las variables ν son condicionalmente independientes o localmente independientes.

La primera parte de la definición de los modelos latentes implica que el área de integración d -dimensional R viene dado por el producto de d intervalos $R_i = (-\infty, \infty)$. La segunda parte de la definición implica que podemos escribir: $g(\nu = u | \eta = h) = \prod_{i=1}^n g(\nu_i = u_i | \eta = h)$ luego:

$$P(\nu = u) = \int \dots \int_{R^d} g(\nu = u | \eta = h) f(h) dh = \int \dots \int_{R^d} \prod_{i=1}^n g(\nu_i = u_i | \eta = h) f(h) dh$$

En la ecuación anterior es necesario especificar lo siguiente:

a) la dimensionalidad del vector de variables latentes;

b) la expresión de la función $g(\nu = u | \eta = h)$;

c) la expresión de la función de densidad $f(h)$.

La finalidad última de los modelos de rasgos latentes es la de ofrecer una representación parsimoniosa de datos multivariados. Es decir, estamos interesados en hallar un modelo del tipo anterior que nos proporcione una adecuada representación de los datos en tan pocas dimensiones como sea posible. Además deseamos que d sea menor que n , a ser posible mucho menor que n .

Los modelos de rasgos latentes son un conjunto amplio de modelos entre los que se encuentran:

a) el modelo factorial, que es un modelo de rasgos latentes caracterizado por asumir que las variables observadas son continuas, que las funciones $g(\nu = u | \eta = h)$ son lineales y que la función de densidad de las variables latentes $f(h)$ es multivariada normal; y

b) los modelos de respuesta a los ítems que son aquellos modelos en los que las demás funciones $g(\nu = u | \eta = h)$ son no lineales.

Como vimos en los apartados anteriores existe una amplia literatura sobre modelos unidimensionales de respuesta a los ítems, es decir, con una única variable latente, $d = 1$.

Se han propuesto modelos multidimensionales para datos dicotómicos y para datos policotómicos ordenados y se dispone de métodos de estimación implementados en programas comerciales para este tipo de datos. Se han propuesto modelos multidimensionales para datos politómicos no ordenados (Takane y de Leeuw, 1987).

Zhang (2004a) probó que cuando se tiene una estructura simple, entonces la estimación por máxima verosimilitud conjunta de los modelos unidimensionales o multidimensionales dan resultados idénticos. Esto no ocurre si se usa máxima verosimilitud marginal. También Zhang (2005b) usó un algoritmo genético en el paso de maximización del algoritmo EM para la estimación de los parámetros cuando hay una estructura compleja.

Haberman, von Davier, y Yi-Hsuan Lee (2008) mostraron empíricamente que los modelos multidimensionales pueden estar basados en distribuciones de la habilidad multivariadas o polatómicas. Estos modelos basados en distintos tipos de distribuciones son comparables cuando hay una estructura simple en términos de ajuste, tiempo de cálculo y estimación de los parámetros.

Rijmen (2009b y 2010) estudió tres modelos multivariados para test basados en testlets, probando que el modelo multidimensional de segundo orden es equivalente al modelo bifactor para testlet. Rijmen et al. (2013) describen cómo los modelos multidimensionales pueden ser entendidos como casos particulares de los modelos lineales generalizados y los modelos mixtos no lineales.

Haberman y Sinharay (2010a y 2010b) estudiaron el uso de modelos multidimensionales para calcular subpuntajes proponiendo estadísticos para determinar cuando los modelos multidimensionales de sub-puntajes agregan valor al puntaje total y a los sub-puntajes basados en la teoría clásica. Probaron que estos nuevos métodos son más exactos que los basados en teoría clásica. Rose, von Davier y Xu (2010) estudiaron, en el contexto de evaluaciones internacionales, las no respuestas con dos modelos multidimensionales. Uno usaba variables indicadoras para determinar cuando las no respuestas ocurrían y el otro realizaba un modelo multidimensional multigrupo.

Van Rijn y Rijmen (2012) explican la paradoja que ocurre en algunos modelos multidimensionales donde responder un ítem adicional correctamente puede resultar en una disminución del puntaje de la variable latente en alguna de las dimensiones. Los autores explican claramente por qué esto ocurre y cómo no sucede en los testlets.

Modelo de respuesta a los ítems multivariado normal para variables dicotómicas

Cuando las variables observadas son dicotómicas, $\nu_i = \{0, 1\}$, y la función que modela la probabilidad de respuesta es la función de distribución de la normal estándar, Φ , las funciones de respuesta a las opciones a un ítem vienen dadas por:

$$P(\nu_i = 1 | \eta = h) = \Phi(\alpha_i + \beta'_i h) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha_i + \beta'_i h} \exp\left(-\frac{y^2}{2}\right) dy$$

$$P(\nu_i = 0 | \eta = h) = 1 - P(\nu_i = 1 | \eta = h),$$

donde, α_i y β'_i son el intercepto y vector de pendientes del ítem i y se asume que los rasgos latentes siguen una distribución multivariada normal. Es decir, $f(h) = \varphi(h) \sim N(0, \Phi)$, donde φ denota una densidad multivariada normal, y Φ es la matriz de correlaciones entre los factores. Dado que podemos escribir:

$$P(\nu_i = u_i | \eta = h) = [P(\nu_i = 1 | \eta = h)]^{\nu_i} [1 - P(\nu_i = 1 | \eta = h)]^{1-\nu_i} \quad \nu_i = \{0, 1\}$$

Sustituyendo obtenemos:

$$P(\nu = u) = \int \dots \int_{R^d} \prod_{i=1}^n [P(\nu_i = 1 | \eta = h)]^{\nu_i} [1 - P(\nu_i = 1 | \eta = h)]^{1-\nu_i} \varphi(h) dh$$

Las ecuaciones anteriores definen el modelo normal para datos dicotómicos.

Existe una forma alternativa de derivar el modelo normal. Considérese un vector de variables \mathbf{y} de dimensiones $n \times 1$ que sigue el modelo del factor común, y asúmase una densidad multivariada normal para y . Por tanto, la densidad marginal de y viene dada por:

$$f^*(y) \sim N(\mu_y, \Sigma_y) = N(0, \Lambda\Phi\Lambda' + \Psi)$$

y la densidad de y condicionada a un valor fijo de η es $g^*(y = y^* | \eta = h) \sim N(\Lambda h, \Psi)$, donde Λ es una matriz de cargas factoriales, Ψ es una matriz diagonal de varianzas residuales y Φ es la matriz de correlaciones entre las variables latentes o factores. Las variables aleatorias y y η no son directamente observables, únicamente observamos

$$\nu_i = \begin{cases} 1 & \text{si } y_i \geq \tau_i \\ 0 & \text{si } y_i < \tau_i \end{cases} \quad i = 1, \dots, n$$

donde el vector de parámetros τ se denomina umbrales. Dado que las variables observables son dicotómicas, las varianzas de las variables y no son identificables, por lo que es habitual fijar su valor igual a la unidad, es decir, $\text{diag}(\Lambda\Phi\Lambda' + \Psi) = I$.

Las ecuaciones anteriores implican que:

$$P(\nu = u) = \int_{R^*} f^*(y) dy$$

donde $f^*(y)$ es la densidad multivariada y R^* es un área n -dimensional de integración dada por el producto de intervalos:

$$R_i^* = \begin{cases} (\tau_i, \infty) & \text{si } \nu_i = 1 \\ (-\infty, \tau_i) & \text{si } \nu_i = 0 \end{cases}$$

El modelo ahora especificado es equivalente al primero.

De igual forma que en el caso unidimensional del anterior modelo podemos deducir el modelo logístico multidimensional quedando:

$$P(U_{ij} = 1 | a_i, d_i, c_i, \theta_j) = c_i + (1 - c_i) \frac{e^{(a_i^t \theta_j + d_i)}}{1 + e^{(a_i^t \theta_j + d_i)}}$$

donde t indica el vector traspuesto.

1.8. Otros modelos

Veremos tres modelos adicionales: el modelo logístico multidimensional, el modelo lineal logístico y el modelo multicomponente. El primer tipo de modelo multidimensional se basa en la función logística acumulada y son generalizaciones de los modelos logísticos unidimensionales donde se reemplaza los parámetros y habilidad por vectores. El modelo logístico toma la forma

$$P(u_i = 1 | \theta_p) = \frac{e^{a_i^t \theta_p + d_i}}{1 + e^{a_i^t \theta_p + d_i}}$$

donde a_i es un vector de igual dimensión que el espacio latente que representan tasas de cambio en cada dimensión y d_i es un escalar relativo a la dificultad del ítem.

Hay dos variantes del modelo como en el caso unidimensional; una variante que considera a todos los a_i iguales (en general valen 1), es decir, es un modelo de la familia de Rasch. La segunda variante considera la posibilidad de aciertos al azar como el modelo 3P.

También existe un modelo parcialmente compensatorio que toma la forma

$$P(u_i = 1 | \theta_p) = c_i + (1 - c_i) \prod_{d=1}^D \frac{e^{a_{id}(\theta_{pd} - b_{id})}}{1 + e^{a_{id}(\theta_{pd} - b_{id})}}$$

donde D es la dimensión del espacio latente, b_{id} indica la dificultad de la tarea en la dimensión d . Es claro que este modelo se puede simplificar tomando los $c_i = 0$ y luego aun más tomando

los $a_i = 1$. Fischer(1973) propone el modelo lineal logístico debido a que, para algunos test, la dificultad del ítem se puede considerar como función de ciertas operaciones cognitivas que intervienen en el proceso de solución del ítem.

En este tipo de modelos cada operación cognitiva tiene asociado un parámetro η_k que refiere a la dificultad de realizar ese proceso cognitivo. También toma en cuenta la cantidad de veces (q_{ik}) que cada operación cognitiva se debe llevar a cabo para ejecutar correctamente la tarea. Por lo tanto, la dificultad del ítem es igual a la suma de todos los parámetros de las operaciones cognitivas ponderados por la frecuencia de su aparición. El modelo logístico lineal puede usarse también para investigar los efectos de la posición de los ítems (Debeer y Janssen, 2012; Fischer, 1995b; Hahne, 2008; Hohensinn et al., 2008; Kubinger, 2008, 2009), efectos de los formatos de respuesta (Kubinger y Draxler, 2006), o efectos de los evaluadores (Draney y Wilson, 2008).

El modelo logístico lineal asume el modelo de Rasch y reemplaza la dificultad b_i por una combinación lineal de K características del ítem. Entonces el modelo expresado como un logit toma la forma $\text{logit}(P(U_i = 1)) = \theta_p - \sum_{k=1}^K q_{ik}\eta_k$ donde los llamados parámetros básicos del modelo (Fischer 1983, 1995b) q_{ik} y η_k son los ponderadores a priori en el ítem de la k -ésima característica y el valor de la k -ésima característica respectivamente. Embretson (1999) propone una combinación del modelo 2P y el lineal logístico al que llama modelo 2P restringido que toma la forma:

$$\text{logit}(P(U_i = 1)) = (\sum_{k=1}^K q_{ik}\tau_k)(\theta_p - \sum_{k=1}^K q_{ik}\eta_k)$$

donde τ_k refiere al impacto de la característica k del ítem i en la discriminación del ítem.

Janssen et al. (2004), De Boeck, (2008) y Mislevy (1988) desarrollan un modelo con efectos aleatorios, donde la dificultad del ítem no está completamente determinada por la estructura lineal de los parámetros básicos, incluyendo un error aleatorio distribuido normalmente en el modelo. Su formulación es:

$$\text{logit}(P(U_i = 1)) = \theta_p - \sum_{k=1}^K q_{ik}\eta_k + \varepsilon_i$$

Rijmen y De Boeck (2002) desarrollaron el modelo lineal logístico con ponderadores aleatorios. Este modelo divide las características de los ítems en dos grupos: las que tienen efectos fijos (η_k igual para todos los sujetos) y las que tienen efectos aleatorios (η_{kp} es específico para cada persona). El modelo queda:

$$\text{logit}(P(U_i = 1)) = \theta_p - \sum_{k=1}^{K_1} q_{ik}\eta_k - \sum_{k=K_1+1}^K q_{ik}\eta_{kp}$$

donde K_1 es el número de efectos fijos y $K_2 = K - K_1$ es el número de efectos aleatorios.

Xie y Wilson (2008) proponen una extensión del modelo logístico lineal donde los efectos de las características de los ítems pueden ser distintos para diferentes grupos de sujetos. Ellos llaman a este hecho funcionamiento diferencial de las facetas, donde faceta refiere a un conjunto de ítems que conforman una característica en común.

Identificar una teoría cognitiva plausible para un test tiene varias ventajas. En principio provee evidencia para la validez de constructo, por otra parte aportar sobre los procesos cognitivos es relevante para el diseño de los test (Henson y Douglas, 2005). Bejar (2010) hace una revisión de las investigaciones que tratan sobre este punto. También es indispensable si se quiere construir ítems automáticamente (Bejar, 2010; Gierl y Haladyna, 2012). Además la evaluación de las diferencias individuales en la habilidades cognitivas son útiles para una interpretación de los puntajes de los test (Hensen et al., 2009; von Davier, 2008a, Leighton y Gierl, 2007; Rupp et al., 2010). Los modelos de múltiples componentes son modelos que relacionan las teorías cognitivas con las propiedades psicométricas de los ítems y las medidas de las personas.

Los modelos multicomponentes más conocidos son el modelo multicomponente de rasgo latente presentado por Whitely (1980) el modelo general de componentes (Embretson, 1984)

y el modelo de rasgo latente multicomponente para diagnóstico (Embretson y Yang, 2013). Omitiremos la formulación matemática de estos modelos.

Yan, Almond, y Mislevy (2004) usaron datos de sustracción mixta y la investigación cognitiva desarrollada por Tatsuoaka para comparar varios modelos de diagnóstico e identifican los tipos de problemas para los cuales son adecuados.

von Davier (2005 y 2008a) presentó una familia de modelos de diagnóstico generales que puede aplicarse tanto a ítems dicotómicos como politómicos y a modelos compensatorios y no compensatorios. Estos modelos se pueden estimar mediante máxima verosimilitud marginal. Esta familia de modelos contiene a los modelos de Rasch unidimensionales y multidimensionales, logístico de dos parámetros, modelo de crédito parcial, modelo de crédito parcial generalizado, modelo de facetas y modelos de perfil. Además, von Davier (2007a, 2007b, 2008b) extendió, en primer lugar, el modelo al caso jerárquico y, en segundo término, al caso mixto, que permite estimar los modelos de diagnóstico en varias poblaciones conocidas, desconocidas y en mezcla de poblaciones no observadas.

Xu (2007) estudió la monotonía en el modelo general de diagnóstico y encontró que, al igual que en el modelo de crédito parcial generalizado, la monotonía se cumple cuando los parámetros de pendiente se restringen a ser iguales. Xu y M. von Davier (2008a) extendieron este modelo usando un modelo log-lineal para reducir el número de parámetros a ser estimados en la distribución de habilidades latentes.

Sorrel et al. (2016), García, Olea y de la Torre (2014) utilizan modelos de diagnóstico cognitivo en los datos obtenidos de aplicar el *situational judgment test* mostrando que el modelo de diagnóstico cognitivo les fue útil para explorar la naturaleza de los constructos subyacentes.

von Davier y von Davier (2004/2007, 2010) presentan una teoría unificada basada en teoría de respuesta al ítem de transformación y equiparación de las escalas. En su trabajo, todo proceso de equiparación es visto como una restricción en el espacio de parámetros y luego se reescribe la función del logaritmo de la verosimilitud y se maximiza bajo restricciones lineales y no lineales con lo cual se logra la equiparación. Xu y M. von Davier (2008c y 2008d) desarrollan para el modelo de diagnóstico general un procedimiento de equiparación.

Wang, Bradlow y Wainer (2002) extienden usando métodos bayesianos los modelos de testlets a test que tienen ítems que puntúan en forma politómica y/o en forma dicotómica. Usan datos que provienen del Test of Spoken English y del test North Carolina Test of Computer Skills.

Sinharay, et al. (2003) utilizan un modelo bayesiano jerárquico para estudiar familias de ítems que toma en cuenta la dependencia a la estructura de la familia y que permite la calibración de la familia más que los ítems individuales. Introducen la llamada función de respuesta esperada de la familia que da la probabilidad de responder correctamente un ítem generado aleatoriamente a partir de la familia.

1.9. Modelos de desdoblamiento

Los modelos de desdoblamiento son modelos de proximidad, donde las categorías de respuesta tienen más probabilidad de ser elegidas cuando la distancia entre los parámetros de los sujetos y la posición del ítem es menor. Los modelos de desdoblamiento fueron inicialmente pensados para medidas de actitudes.

Se denomina modelos de desdoblamiento a los modelos de TRI para ítems de actitudes o preferencias de respuesta directa y que tienen funciones de respuesta unimodales, más que monótonas (Sijtsma y Junker, 2006).

Tienen su origen en la versión probabilística de los modelos de desdoblamiento de Coombs ya mencionados. Sin embargo, es en época reciente que estos modelos unimodales y los modelos de TRI monótonos convencionales se conectaron a través del proceso de datos perdidos (Andrich y Luo, 2003 y Verhelst y Verstralen, 1993 en Sijtsma y Junker, 2006).

Varios modelos de desdoblamiento a ítems unidimensionales se han desarrollado para medidas de actitudes; algunos adecuados para respuesta binaria y otros apropiados para respuestas graduadas. Se destacan el Modelo Parella, el modelo graduado generalizado de desdoblamiento y el modelo de coseno hiperbólico. Por ejemplo, el modelo del coseno hiperbólico para respuestas de actitud dicotómicas ($X_j = 1$ para acuerdo ; $X_j = 0$ para desacuerdo) formulado por:

$$P_j(\theta) = P[X_j = 1|\theta] = \frac{e^{\gamma_j}}{e^{\gamma_j} + \cosh(\theta - \beta_j)} = \frac{\exp\{\theta - \beta_j + \gamma_j\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}}$$

en la cual β_j es la posición de las personas en la escala de preferencia o de actitud que más probablemente aprueban el ítem j , y γ_j es el máximo logaritmo de la razón de probabilidades de la aprobación del ítem. Este modelo puede verse como un modelo de crédito parcial (Sijtsma y Junker, 2006:92) con:

$$\begin{aligned} R_{j0}(\theta) &= P[\xi_j = 0|\theta] = \frac{1}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \\ R_{j1}(\theta) &= P[\xi_j = 1|\theta] = \frac{\exp\{\theta - \beta_j + \gamma_j\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \\ R_{j2}(\theta) &= P[\xi_j = 2|\theta] = \frac{\exp\{2(\theta - \beta_j)\}}{1 + \exp\{\theta - \beta_j + \gamma_j\} + \exp\{2(\theta - \beta_j)\}} \end{aligned}$$

donde los datos completos ξ_j están codificados como $\xi_j = 0$ si $\theta - \beta_j \ll 0$; $\xi_j = 1$ si $\theta - \beta_j \approx 0$; $\xi_j = 2$ si $\theta - \beta_j \gg 0$

Otro modelo de la familia paramétrica ha sido desarrollado por Roberts, Donoghue y Laughlin (2000) y Johnson y Junker (2003) que generalizaron la idea de los datos perdidos y usaron estimación Bayesiana para el desarrollo de una familia de modelos de desdoblamiento paramétricos de TRI.

Post (1992) desarrolló un método de escalamiento no paramétrico con modelos de desdoblamiento basado en la probabilidad de desigualdades, orden estocástico e ideas relacionadas. La idea clave es que $P[X_i = 1|X_j = 1]$ crecerá cuando las modas de las funciones de respuesta al ítem para ítems i y j estén más cercanas. Esta propiedad es llamada “unimodalidad manifiesta”. Post mostró que esta propiedad sigue de las propiedades de orden estocástico de la función de respuesta al ítem. Johnson (2006) reexaminó el enfoque de Post y lo conectó con la teoría de la estimación no paramétrica para modelos de desdoblamiento basados en los trabajos de Stout (1990), Ramsay (1991) y Hemker et al. (1997). Por ejemplo, Sijtsma y Junker (2006) apuntan que Johnson estableció las propiedades de monotonicidad y consistencia de las puntuaciones de Thurstone bajo un conjunto de supuestos similares a los de Hemker et al. (1997) y Stout (1990) y la estimación de la función de respuesta al ítem mediante regresión no paramétrica sobre los puntajes de Thurstone, similar al enfoque de Ramsay (1991).

Capítulo 2

Modelos no paramétricos de teoría de respuesta al ítem

2.1. Introducción

Se comenzará brindando un panorama general sobre los modelos no paramétricos desde su inicio a la actualidad. Mokken en 1971 publica su libro *A Theory and Procedure of Scale Analysis* donde presenta dos modelos no paramétricos de TRI: el modelo monótono homogéneo y el modelo de doble monotonidad.

La TRI no paramétrica relaja los supuestos de independencia local, monotonía y unidimensionalidad, en tanto mantiene importantes propiedades métricas, tales como la escala ordinal de las personas (Junker, 2001; Stout, 1990, 2002). Es claro que estos modelos de TRI al ser no paramétricos se abstienen de una definición paramétrica de la función de respuesta, como lo hacen los modelos de TRI hasta ahora discutidos en esta tesis. Esto hace que sean menos restrictivos y más flexibles. Las curvas de respuestas están sujetas a restricciones de orden, por ejemplo, la monotonía de las CCI.

Uno de los principales cuestionamientos que se pueden hacer al enfoque paramétrico, es que la dependencia entre la probabilidad y el rasgo es forzada a ser de determinada forma funcional y no se toma en cuenta hechos tales como la no monotonicidad de los ítems o desviaciones del formato impuesto. También está de por medio la complejidad de los cálculos, tanto si se emplea el paradigma de Birnbaum (máxima verosimilitud marginal con el algoritmo EM) como procedimientos bayesianos. En los algoritmos usuales como el Logist (Wingersky, Patrick y Lord, 1988) y Bilog (Mislevy y Bock 1982) la velocidad de convergencia es lineal y usualmente lenta.

Mokken (1971) critica la utilización de modelos paramétricos afirmando que para hacer uso de ellos se necesita un amplio conocimiento de los ítems y los sujetos examinados y que este conocimiento es asequible cuando se tiene acceso a una gran cantidad de ítems y examinados y existen áreas en que por razones prácticas nunca se lograría. Así busca una alternativa basada solo en restricciones de orden.

Wahba (1990) propone un modelo parcialmente no paramétrico. Otros trabajos que apuntan en la misma dirección son los de Levine (1984, 1985), Drasgow, Levine, Williams, Mc Laughlin y Candell (1989), Ramsay y Abrahamowicz (1989), Ramsay y Winsberg (1991).

Desde el punto de vista no paramétrico tenemos el modelo homogéneo monótono y el modelo homogéneo de doble monotonicidad. Para ver un desarrollo detallado de estos modelos para ítems dicotómicos y su extensión a politómicos se puede consultar el libro de Sijtsma y Molenaar (2002) o los capítulos de Mokken (1997) y Molenaar (1997).

Estudiaremos en este capítulo los modelos de Mokken, modelos bayesianos no paramétricos, modelos basados en regresión no paramétrica con núcleos y un enfoque desde el análisis de datos funcionales.

2.2. Modelos de Mokken

Las primeras referencias a estos modelos se encuentran en Mokken (1971), Henning (1976), Niemöller y Van Schuur (1983), Mokken y Lewis (1982), Sijtsma (1988) y Giampaglia (1990). Más recientemente hay trabajos de Mokken (1997), Sijtsma (1998, 2001), Molenaar y Sijtsma (2000), Junker (2001) y Junker y Sijtsma (2001). El tratamiento matemático del modelo de monotonía homogénea lo podemos encontrar en Holland y Rosenbaum (1986), Holland (1990), Stout (1987, 1990), Junker (1993) y Ellis y Junker (1997).

Una comparación entre los modelos paramétricos y no paramétricos se encuentra en Meijer, Sijtsma y Smid (1990), y en De Koning, Sijtsma y Hamers (2002).

Douglas, Kim, Habing y Gao (1998) dan métodos para investigar la independencia local, en tanto métodos para investigar CCI no decrecientes desde un punto de vista práctico son presentados en Ramsay (1991), Molenaar y Sijtsma (2000) y Douglas y Cohen (2001).

Consideremos una muestra aleatoria de N sujetos tomados de una población que responden a n ítems dicotómicos.

Se supondrá que se cumplen los supuestos de unidimensionalidad, independencia local y que las CCI cumplen la propiedad de monotonía homogénea respecto al rasgo, es decir, que para todo i y rasgos θ_l y θ_m entonces $\theta_l \leq \theta_m$ si y solo si $P_i(\theta_l) \leq P_i(\theta_m)$.

Esta propiedad es equivalente a la llamada propiedad de orden similar que establece que un conjunto de CCI ordena similarmente a un conjunto de N sujetos con rasgos $\theta_1, \theta_2, \dots, \theta_N$ si para algún i $P_i(\theta_1) \leq P_i(\theta_2) \leq \dots \leq P_i(\theta_N)$ entonces para todo ítem k $P_k(\theta_1) \leq P_k(\theta_2) \leq \dots \leq P_k(\theta_N)$.

Estas propiedades son suficientes para tener un orden conjunto de los sujetos respecto a un rasgo latente, sin embargo, no son suficientes para un orden uniforme de las dificultades de los ítems. Para esto se requerirá una nueva condición llamada de doble monotonía o monotonía homogénea fuerte.

Formalmente, un conjunto de ítems homogéneamente monótonos satisfacen la propiedad de doble monotonía respecto a θ , si para todo par de ítems i y j que para algún θ_0 se cumple que $P_i(\theta_0) \leq P_j(\theta_0)$, entonces para todo θ $P_i(\theta) \leq P_j(\theta)$.

Es decir, las n curvas no se intersectan y las CCI se pueden ordenar de tal manera que $P_1(\theta) \leq P_2(\theta) \leq \dots \leq P_n(\theta)$ para todo θ . Esto nos permite ordenar los ítems según dificultad.

2.2.1. Modelo monótono homogéneo

Este modelo se soporta en las hipótesis de independencia local y monotonía homogénea. Sea $X_+ = \sum_{i=1}^n X_i$ el puntaje total obtenido en el test; los sujetos son ordenados mediante este estadístico.

Este modelo mide los sujetos en una escala ordinal formalmente, sea ζ un valor arbitrario del rasgo y sean $0 \leq s < t \leq n$ dos posibles valores del puntaje total entonces $P(\theta > \zeta | X_+ = s) \leq P(\theta > \zeta | X_+ = t)$. En otras palabras tiene un orden estocástico de la variable latente θ por medio del puntaje total.

De lo anterior se deduce que para $0 \leq s < t \leq n$, $E(\theta|X_+ = s) \leq E(\theta|X_+ = t)$ la cual implica que la media del rasgo en el subgrupo de individuos con puntaje t es mayor a la media del rasgo en el subgrupo que tiene puntaje s si $0 \leq s < t \leq n$. Esto justifica el uso de $X_+ = \sum_{i=1}^n X_i$ para ordenar a los sujetos y además se deriva una escala ordinal para la medición de las personas. También se cumple para $\theta_1 < \theta_2$ que: $P(X_+ > x|\theta = \theta_1) \leq P(X_+ > x|\theta = \theta_2)$ y $E(X_+|\theta = \theta_1) \leq E(X_+|\theta = \theta_2)$

Teorema 2.2.1. (Mokken 1971) Si se cumple la propiedad de monotonía homogénea y la independencia local entonces la covarianza entre todos los pares de ítems es no negativa.

Demostración Sea $F(\theta)$ la función de distribución del rasgo latente y consideremos dos ítems cualesquiera i, j . Se tiene que: $P(X_i = 1) = \int P_i(\theta) dF(\theta)$ y $P(X_i = 1, X_j = 1) = \int P_i(\theta) P_j(\theta) dF(\theta)$.

Debido a que se trata de ítems dicotómicos, se cumple que:

$$\text{Cov}(X_i, X_j) = P(X_i = 1, X_j = 1) - P(X_i = 1)P(X_j = 1) = \iint P_i(\theta)P_j(\xi) dF(\theta) dF(\xi) - \int P_i(\theta) dF(\theta) \int P_j(\xi) dF(\xi) = \frac{1}{2} \iint [P_i(\theta) - P_i(\xi)] [P_j(\theta) - P_j(\xi)] dF(\theta) dF(\xi)$$

En la región donde $\theta > \xi$ ambos términos son no negativos debido a que las CCI son monótonas y en donde $\theta < \xi$ ambos son no positivos, luego la covarianza es no negativa.

2.2.2. Modelo de doble monotonicidad

Este modelo incorpora adicionalmente la hipótesis de no intersección de las CCI. Es decir, para todo par de ítems i, j si sabemos que existe un valor θ_0 para el cual $P(X_i = 1|\theta_0) \leq P(X_j = 1|\theta_0)$, entonces para todo valor de θ se tiene que:

$$P(X_i = 1|\theta) \leq P(X_j = 1|\theta)$$

Sijtsma y Junker (2006) definen que n funciones de respuesta tienen un orden invariante si ellas no se intersectan permitiendo que para ciertos θ (o intervalos) sean iguales.

En el caso de ítems dicotómicos tenemos que $E(X_i|\theta) = P_i(\theta)$. A partir de esto se obtiene, ya que las curvas no se cortan que:

$$E(X_1|\theta) \leq E(X_2|\theta) \leq \dots \leq E(X_n|\theta)$$

para todo θ . Se dice cuando esto ocurre que los ítems tienen un orden invariante.

Ligtvoet et al. (2010, 2011) y Tijmstra et al. (2011) dan resultados y métodos para estudiar y evaluar el orden invariante de los ítems. Para poder establecer cuál es el orden de precedencia entre los ítems, ya que no es posible observar directamente las esperanzas condicionales, se recurre a la proporción de respuestas correctas. Como en nuestras suposiciones hemos considerado que el rasgo latente tiene densidad $f(\theta)$, entonces la proporción de respuestas correctas en el ítem i puede determinarse como:

$$P_i = \int_{-\infty}^{+\infty} P_i(\theta) f(\theta) d\theta = \int_{-\infty}^{+\infty} E(X_i|\theta) f(\theta) d\theta$$

Si el ítem i tiene menor o igual proporción de respuestas correctas que el ítem j entonces,

$$P_i - P_j = \int_{-\infty}^{+\infty} [P_i(\theta) - P_j(\theta)] f(\theta) d\theta = \int_{-\infty}^{+\infty} [E(X_i|\theta) - E(X_j|\theta)] f(\theta) d\theta \leq 0 \text{ de donde se}$$

deduce que si $P_i \leq P_j$ y sabemos que los ítems tienen la propiedad de orden invariante, entonces, $E(X_i|\theta) \leq E(X_j|\theta)$ para todo θ . Vemos así que hay dos propiedades de monotonía: una sobre las CCI y otra dada por el orden invariante.

No se realizan asunciones sobre la densidad $f(\theta)$, esto implica que si tenemos dos subpoblaciones con densidades $f_1(\theta)$ y $f_2(\theta)$ el orden de la proporción de respuestas correctas en cada subpoblación es el mismo, lo cual es un resultado relevante.

Mokken (1971) prueba que para todo ítem i y todo par de ítems j y k con $j < k$ se cumple:

$$\begin{aligned} P(X_i = 1, X_j = 1) &\leq P(X_i = 1, X_k = 1) \\ P(X_i = 0, X_k = 0) &\leq P(X_i = 0, X_j = 0) \end{aligned}$$

Rosenbaum (1987) probó que si i es más difícil que k entonces $P(U_i = 1 | X_+(u_2)) \leq P(U_k = 1 | X_+(u_2))$, donde u_2 es el vector de respuestas en los $n-2$ ítems restantes.

Molenaar (1997) generalizó los dos modelos de Mokken al caso politómico con ítems con más de dos puntajes ordenados. Estas extensiones no implican el orden estocástico de la variable latente por el puntaje total (Hemker et al., 1997). Aunque, Van der Ark (2005), basado en simulaciones de gran cantidad de datos, muestra que el orden estocástico por el puntaje total en el caso del modelo homogéneo ocurre aproximadamente. Van der Ark y Bergsma (2010) probaron que para el modelo de monotonía homogéneo para ítems politómicos se cumple el orden estocástico débil.

2.2.3. Estimación de la habilidad

Para la estimación de las habilidades hay que recurrir a métodos indirectos. Una primera alternativa es utilizar el puntaje total que con base en las observaciones anteriores está correlacionado positivamente con el rasgo y es un estadístico óptimo para la clasificación binaria.

Schriever (1985) propone utilizar análisis de correspondencia múltiple. La primera componente principal de la matriz de correlaciones de los ítems determina la habilidad.

El primer componente puede escribirse como:

$$Y = \sum_{i=1}^n \alpha_i (-1)^{1-u_i} \sqrt{\frac{1-P_i}{P_i}}$$

donde α es el vector propio del valor propio mayor y $\alpha^t \alpha = 1$.

Lewis (1983) propone un método bayesiano que permite obtener intervalos de confianza para el rasgo y puede ser usado en tests adaptativos informatizados.

Coeficiente H de escalabilidad

Loevinger (1947, 1948) propuso el coeficiente H pero no tuvo mayor aplicabilidad hasta que lo adaptó Mokken (1971) para definir una familia de coeficientes que indican la escalabilidad monótona homogénea para:

- a) un par de ítems;
- b) un ítem respecto a los restantes;
- c) el conjunto total de los ítems.

Consideremos dos ítems i, j dicotómicos, al aplicarse los dos ítems sobre N sujetos se obtiene una tabla 2x2 con las posibles salidas. Si mantenemos fijas las marginales podemos obtener la tabla que produce la máxima covarianza entre los ítems, luego el coeficiente H se define como:

$$H_{ij} = \frac{Cov(X_i, X_j)}{Cov_{\max}(X_i, X_j)} = \frac{P_{ij} - P_i P_j}{P_i - P_i P_j} = 1 - \frac{P_i - P_{ij}}{P_i(1 - P_j)}$$

donde se utiliza el hecho que el máximo de la covarianza condicionada a las marginales se da cuando $P_{ij} = P_i$.

Se puede definir H a partir de las probabilidades de error de Guttman.

Consideremos dos ítems: i, j tal que $i < j$, o sea $P_i < P_j$ la probabilidad de error de Guttman se define como $e_{ij} = P(X_i = 1, X_j = 0) = P_i - P_{ij}$, y el valor esperado del error bajo la hipótesis de independencia es $e_{ij}^o = P_i(1 - P_j)$. Luego $H_{ij} = 1 - \frac{e_{ij}}{e_{ij}^o}$.

Valores altos de H_{ij} se corresponden con CCI discriminantes ($e_{ij} \approx 0$) y valores bajos con CCI que no discriminan (son independientes), en tanto valores negativos contradicen la monotonía homogénea.

El coeficiente de escalabilidad del ítem i con el test se define como:

$$H_i = \frac{Cov(X_i, X_+^{-i})}{Cov_{\max}(X_i, X_+^{-i})} = \frac{\sum_{j \neq i} (P_{ij} - P_i P_j)}{\sum_{j > i} (P_i - P_i P_j) + \sum_{j < i} (P_i - P_i P_j)} = 1 - \frac{\sum_{j \neq i} (P_i - P_{ij})}{\sum_{j > i} (P_i - P_i P_j) + \sum_{j < i} (P_i - P_i P_j)}$$

Finalmente, el coeficiente de escalabilidad del test es:

$$H = \frac{\sum_i Cov(X_i, X_+^{-i})}{\sum_i Cov_{\max}(X_i, X_+^{-i})} = 1 - \frac{\sum_i \sum_{j \neq i} (P_i - P_{ij})}{\sum_i \sum_{j > i} (P_i - P_i P_j) + \sum_i \sum_{j < i} (P_i - P_i P_j)}$$

Se puede ver fácilmente que:

$$\begin{aligned} \min_j H_{ij} &\leq H_i \leq \max_j H_{ij} \\ \min_i H_i &\leq H \leq \max_i H_i \\ \min_{i,j} H_{ij} &\leq H \leq \max_{i,j} H_{ij} \end{aligned}$$

$$0 \leq H_{ij} \leq 1 \quad 0 \leq H_i \leq 1 \quad 0 \leq H \leq 1$$

Mokken (1971) propone que una escala es útil si H es mayor o igual a una constante, que debe ser al menos mayor o igual a 0.3 y clasifica las escalas como débiles si $0.3 \leq H < 0.4$ moderadas si $0.4 \leq H < 0.5$ y fuertes si $H \geq 0.5$.

Van der Ark et al. (2008) usó modelos marginales para estimar los errores estándar de los coeficientes de escalabilidad estimados y Kuijpers (2013) extendió estos resultados a ítems politómicos.

Rosenbaum (1984) y luego Holland y Rosenbaum (1986) probaron el siguiente resultado (asociación condicional) para los dos modelos. Si partimos el vector de respuestas u en dos subconjuntos u_1 y u_2 y consideremos f, g dos funciones monótonas no decrecientes definidas sobre u_1 y h una función definida sobre u_2 , entonces

$$Cov(f(U_1), g(U_1)|h(U_2)) \geq 0$$

También se puede probar que:

a) Todo par de ítems i y j están condicionalmente asociados en todo subconjunto de sujetos con un puntaje total t en los restantes ítems.

b) Para todo ítem i , P_i es no decreciente si se toma sobre los subgrupos de sujetos ordenados en forma creciente en función del puntaje total sobre los $n-1$ ítems restantes.

c) Lord y Novick, (1968) prueban que la esperanza condicional (respecto de θ) del puntaje total es una función monótona no decreciente.

d) Grayson (1988) mostró que el puntaje total (condicionado a θ) tiene verosimilitud monótona, o sea, $\frac{P(X_+ = t|\theta)}{P(X_+ = s|\theta)}$ $0 \leq s < t \leq n$ es no decreciente como función de θ , es decir, $X_+|\theta$ es un estadístico óptimo para la clasificación binaria.

Un caso simple ocurre cuando tomamos $f(U_1) = U_i$ y $g(U_1) = U_j$ y no condicionado obtenemos:

$$Cov(U_i, U_j) \geq 0$$

para todo i, j con $i \neq j$

Mokken (1971) utiliza este hecho para definir lo que se entiende por escala.

Definición 2.2.2. Un conjunto de n ítems constituyen una escala al nivel c (Escala de Mokken) si $\rho_{ij} > 0$ para todo i, j con $i \neq j$ y $H_i \geq c$ para todo i para algún valor $0 < c < 1$.

Si en la asociación condicional tomamos $h(U_2) = U_k$ con f y g como antes entonces,

$$Cov(U_i, U_j | U_k = a_k) \geq 0$$

para todo i, j con $i \neq j$ y $k \neq i, j$.

Si se define el puntaje restante como $h(U_2) = S_{(ij)} = \sum_{k \neq i, j} U_k$ entonces,

$$Cov(U_i, U_j | S_{(ij)} = s) \geq 0$$

para todo i, j con $i \neq j$ y $s = 0, \dots, n - 2$.

Stout (2002) y Van Abswoude et al. (2004) utilizan varios métodos basados en cluster para investigar la independencia local y la unidimensionalidad. Straat et al. (2016) usa las ecuaciones precedentes para decidir cuando un conjunto de ítems son localmente independientes, identificando los ítems que con mayor frecuencia están envueltos en dependencias locales con otros ítems.

2.3. Estimación bayesiana no paramétrica

Karabatsos (2016) presenta para modelos infinitos mixtos de TRI un enfoque basado en regresión no paramétrica bayesiana, técnica desarrollada por Karabatsos y Walker (2012a) y para el que también desarrolló un software (Karabatsos 2014a, b). Notemos por U un vector de dimensión n que representa las respuestas a los n ítems de un test, u representará una realización de U . Se asume también que cada ítem i puede tomar $m_i + 1$ valores $0, 1, \dots, m_i$.

Definición 2.3.1. Un modelo de TRI mixto discreto toma la forma general:

$$(2.1) \quad f_{G_x}(u|x) = \int f(u|x; \beta, \Psi(x)) dG_x(\Psi) = \sum_{j=1}^J f(u|x; \beta, \Psi_j(x)) \omega_j(x)$$

donde $f(u|x; \beta, \Psi(x))$ es el núcleo de la mezcla y G_x es la distribución de la mezcla, β es un vector de parámetros fijos, $\Psi_j(x)$ son parámetros aleatorios dependientes de x o no. Los $\omega_j(x)$ son ponderadores que para cada valor fijo de x suman 1.

Como vemos la definición implica que la función de probabilidad es la mezcla de J funciones de probabilidad. El modelo se llamará discreto si G_x lo es y se llamará continuo si G_x lo es. Si la suma es finita se dirá que el modelo es finito y si se tiene una serie se dirá que el modelo es infinito. Un ejemplo simple es el modelo mixto finito de Rasch para ítems dicotómicos (Rost, 1990, 1991). El modelo general de mezcla infinito asume que:

$$(2.2) \quad G_x(\cdot) = \sum_{j=1}^{\infty} \omega_j(x) \delta_{\Psi_j(x)}(\cdot)$$

donde $\delta_{\Psi_j(x)}(\cdot)$ es la distribución degenerada en el soporte de Ψ . Este modelo queda completamente especificado mediante las distribuciones a priori de $\Psi_j(x)$, $\omega_j(x)$ $j = 1, 2, \dots$ y β .

Un ejemplo de esto es el modelo mixto de TRI con el proceso de Dirichlet, en el cual se asume que las distribuciones de mezcla no dependen de x (Sethuraman, 1994). Una generalización de la distribución a priori del proceso de Dirichlet es la a priori de Pitman-Yor (Ishwaran y James, 2001). MacEachern (1999, 2000, 2001) da una generalización para el caso en que las distribuciones de mezcla sean dependientes de la covariable.

El modelo bayesiano no paramétrico de TRI es un caso particular de la regresión bayesiana no paramétrica introducida por Karabatsos y Walker (2012a) quienes demuestran que el modelo tiene mejor funcionamiento predictivo que el proceso de Dirichlet mixto y el proceso de Dirichlet dependiente mixto. Este modelo sirve tanto para ítems dicotómicos como politómicos.

El modelo dicotómico considera P personas y n ítems. Se define el conjunto de los pares $\mathfrak{S} = \{(u_{pi}, x_{pi})\}$ $i = 1, \dots, n$ $p = 1, \dots, P$ donde u indica la respuesta del sujeto p al ítem i y x una variable que describe a la persona p . El modelo toma la forma:

$$(2.3) \quad f(\mathfrak{S}|x; \zeta) = \prod_{p=1}^P \prod_{i=1}^n f(u_{pi}|x_{pi}; \zeta)$$

$$(2.4) \quad f(u_{pi}|x_{pi}; \zeta) = P(U_{pi} = 1|x_{pi}; \zeta)^{u_{pi}} (1 - P(U_{pi} = 1|x_{pi}; \zeta))^{1-u_{pi}}$$

$$(2.5) \quad P(U_{pi} = 1|x_{pi}; \zeta) = \int_0^\infty \sum_{j=-\infty}^{+\infty} n(z|\mu_j x_{pi}^t \beta; \sigma^2) \omega_j(x_{pi}; \beta_\omega; \sigma_\omega) dz$$

$$(2.6) \quad \omega_j(x; \beta_\omega; \sigma_\omega) = \Phi\left(\frac{j - x^t \beta_\omega}{\sigma_\omega}\right) - \Phi\left(\frac{j - 1 - x^t \beta_\omega}{\sigma_\omega}\right)$$

Para la estimación se utiliza un método Monte Carlo basado en cadenas de Markov. Utilizando Gibbs sampling para modelos Bayesianos de mezcla infinita, Kalli et al. (2011) mostraron que es posible muestrear la distribución posterior del modelo bayesiano condicionada a las variables latentes mediante un método Monte Carlo basado en cadenas de Markov. Robert y Casella, (2004) muestran que si la a priori considerada es propia, entonces el proceso de muestreo conduce asintóticamente a una cadena de Markov de Harris de tiempo discreto, por lo cual, luego de marginalizar en las variables latentes, muestra que la distribución a posteriori es estacionaria. Se pueden ver los detalles de este proceso en Karabatsos y Walker (2012a). El resultado es asintótico, pero es claro que solo puede considerarse el método Monte Carlo finito. La convergencia a la muestra de las distribuciones a priori pueden obtenerse a partir de dos procedimientos detallados en Geyer, (2011). Karabatsos (2016) ilustra el procedimiento a partir de datos reales con ítems politómicos.

Como parte de las ventajas de este modelo es que muestra estimadores robustos de la habilidad de las personas y de la dificultad de los ítems y tiene muy buena performance predictiva para datos sin outliers. Otra virtud es la existencia de un software libre para la estimación *Bayesian Regression: Nonparametric and Parametric Models* (Karabatsos, 2014, 2015, 2016).

2.3.1. Estimación usando deconvolución

Ferrando (2004) utiliza un enfoque basado en regresión no paramétrica con errores en la variables para ítems de personalidad de respuesta continua. Fan y Truong (1993) plantean un método para cuando las variables independientes pueden no ser directamente observadas. Ferrando (2004) aplica a la estimación de las funciones características del ítem el modelo de Fan y Truong el cual toma la forma. Sea $\hat{\theta}_i = \theta_i + \varepsilon_i$ un estimador de la habilidad donde ε_i es el error

de estimación. Fan y Truong (1993) suponen que los errores ε tiene una distribución conocida, y que ε es independiente de θ y de la respuesta a los ítems X. Si se cumple lo anterior recordando que la distribución de la suma es la convolución de las distribuciones de los sumandos, se obtiene que la función de densidad de $\hat{\theta}$ viene dada por

$$f_{\hat{\theta}}(\hat{\theta}) = \int_{-\infty}^{\infty} f_{\theta}(\hat{\theta} - \theta) dF_{\varepsilon}(\theta)$$

donde $F_{\varepsilon}(\cdot)$ es la función de distribución de ε . Luego la densidad marginal de θ puede ser estimada usando el procedimientos de deconvolución. Fan y Truong(1993) hallaron un estimador de $f(\theta)$ obtenido por deconvolución y que puede ser escrito como un estimador de regresión no paramétrica donde el núcleo es

$$K_n(X) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\phi_K(t)}{\phi_{\varepsilon}(\frac{t}{h_n})} dt$$

siendo $\phi_K(\cdot)$ la transformada de Fourier del núcleo original, que se usaría para estimar la regresión en caso que la variables fuese medida sin error y $\phi_{\varepsilon}(\cdot)$ la función característica del error. Fan y Truong (1993) propusieron el siguiente estimador regresión:

$$\hat{m}_n(\theta) = E(X_j|\theta) = \frac{\sum_{i=1}^N K_n(\frac{\theta - \hat{\theta}_i}{h_n}) X_{ij}}{\sum_{i=1}^N K_n(\frac{\theta - \hat{\theta}_i}{h_n})}$$

Es claro que los supuestos de Fan y Truong (1993) no se pueden aplicar exactamente a la estimación de la función característica de los ítems, ya que es claro que los errores dependen de la estimación del rasgo. Para lograr una aproximación satisfactoria a estos supuestos Ferrando (2004) corrige el sesgo de las estimaciones. Sea $\hat{\theta}$ la estimación por máxima verosimilitud de θ obtenida de un conjunto de ítems dicotómicos, además se supone que los ítems siguen un modelo dado (en el caso de Ferrando 2P) y que los parámetros de los ítems son fijos y conocidos (logra esto al ser estimados en una muestra grande). También asume que los errores de medición obtenidos al aplicar el método de máxima verosimilitud anteriormente son independientes de las respuestas observadas en los ítems de respuesta continua.

La distribución condicional del estimador máximo verosimil $\hat{\theta}$ es asintóticamente normal con media θ y varianza $1/I(\theta)$, donde $I(\theta)$ es la función de información del test. Aunque Samejima (1977) mostró empíricamente que esta distribución condicional es aproximadamente normal para un test de longitud finita si la información es suficientemente grande el estimador $\hat{\theta}$ es sesgado para un test de longitud finita. Samejima (1993) estudió la función de sesgo para un caso particular del modelo de 2P y encontró que el sesgo de $\hat{\theta}$ tiende a ser chico si las dificultades de los ítem se distribuyen en un rango amplio, uniformemente, y si además las discriminaciones son bajas.

Supuesto entonces que $\hat{\theta}$ es condicionalmente insesgado, se sigue que θ y ε son no correlacionados aunque no independientes. Como la $I(\theta)$ (por ejemplo en el modelo 2P) depende del nivel de θ es claro que no son independientes. Este problema puede ser resuelto de dos formas: (a) transformando la escala original usando la raíz cuadrada de la función de información o (b) eligiendo un conjunto adecuado de ítems para los cuales la función de información del test es aproximadamente constante en un rango de θ . Samejima (1998) propuso y desarrolló la primera solución. La segunda solución está basada en los resultados que la información del test es la suma de la información de los ítems y fue desarrollado en un procedimiento de cuatro pasos por Lord (1977). Si se logra que la información sea constante entonces ε y θ serán independientes. O sea si la distribución de θ es normal, la distribución marginal de ε se aproximará a ser normal

con media cero y varianza σ_ε^2 . Luego en estas condiciones puede asumirse que los supuestos de Fan y Truong (1993) se cumplen.

Como caso particular Ferrando (2004) usa el núcleo normal entonces la transformada de Fourier de K es $\phi_K(t) = e^{-\frac{1}{2}t^2}$. Asumiendo que la distribución del error es normal con media cero y varianza σ_ε^2 , se sigue que la función característica a usar es $\phi_\varepsilon(t) = e^{-\frac{1}{2}(\sigma_\varepsilon t)^2}$.

Con esto entonces operando se obtiene que $K_n(X) = \frac{1}{c\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x}{c})^2}$ donde $c^2 = (1 - \frac{\sigma_\varepsilon^2}{h_n^2})$.

2.4. Modelo no paramétrico unidimensional usando regresión no paramétrica

Este enfoque del problema en Psicometría fue introducido por Ramsay (1991) y puede ser considerado como una técnica del análisis de datos funcionales. Ramsay (1991) propone estimar el rasgo no paramétricamente y utilizar los métodos de regresión no paramétrica mediante núcleos para obtener una estimación de las CCI.

Este método tiene varias ventajas: la primera de ellas es la no imposición de ninguna restricción sobre la forma funcional de las CCI, y además, el método de estimación es no iterativo, rápido, eficiente y muy fácil de programar. Además, estas técnicas permiten tener herramientas de la TRI para situaciones de muestras de tamaño moderado.

Es de destacar que, para obtener la estimación del rasgo, debe conocerse su distribución, pero si se ignora, lo que se estima es una transformación del rasgo dada por $F^{-1}(\Theta)$ que tiene distribución uniforme en $[0,1]$. Esto no constituye una limitación, pues como lo expresa Ramsay (1991): “*en el contexto del análisis de ítems, un test no puede proporcionarnos más que información sobre el orden de los examinados*”. Esto es conocido como la pérdida de la identificabilidad.

Se consideran N sujetos que responden a n ítems dicotómicos o politómicos y se estima el vector de rasgos θ y las curvas características de los ítems (CCI) mediante estimadores no paramétricos.

Sea $P_{jm}(\theta)$ la probabilidad que un examinado con nivel de rasgo θ seleccione la opción m del ítem j con $j = 1, \dots, n$; $m = 1, \dots, M_j$. Tenemos entonces variables indicadoras Y_{jm}^i , que valen 1 si el sujeto i eligió la opción m en el ítem j , y 0 en caso contrario. Es decir, $P_{jm}(\theta) = P(Y_{jm} = 1 | \Theta = \theta)$.

Debido a que la regresión no paramétrica entre dos variables X e Y supone el conocimiento de ambas, y en nuestro caso se desconoce la variable θ , el primer paso que Ramsay (1991) se plantea es estimar los rasgos θ . Para ello, supone que estas tienen alguna distribución conocida $F(\theta)$.

Consideremos un estadístico T y sean T_i los valores de los sujetos. Sea r_i el rango del i -ésimo individuo. Para cada valor r_i consideremos la quartila q_i respecto a alguna distribución. En general, se usa la normal típica, y si consideramos el rasgo como proporciones, se puede usar la distribución beta, por ejemplo, la distribución $B(2.5, 2.5)$.

Con base en el valor de T_i se ordena a los individuos, obteniéndose ahora los patrones de respuestas ordenados. Se estiman los P_{jm} mediante regresión no paramétrica entre las variables binarias $Y_{jm}^{(i)}$ y el vector de rasgos $\widehat{\theta}_1 = q_1, \dots, \widehat{\theta}_N = q_N$.

Se suele utilizar como estadístico el puntaje total, pero Ramsay (1991) propone un estadístico alternativo.

Sea $W_{jm} = \text{logit}(P_{jm}^{(75)}) - \text{logit}(P_{jm}^{(25)})$ donde $P_{jm}^{(75)}$ y $P_{jm}^{(25)}$ son las proporciones de examinados que caen en el 25 % superior e inferior de la distribución de los puntajes totales de quienes eligen la opción m en el ítem j . Entonces utiliza el estadístico: $T_i = \sum_j \sum_m Y_{jm}^{(i)} W_{jm}$ o sea, T_i es la suma de pesos para las opciones actualmente elegidas. Este estadístico tiene la ventaja de que produce menos empates y usa información sobre la elección de alternativas incorrectas.

El estimador por núcleos de $P(\theta)$ es el promedio ponderado.

$$(2.7) \quad \hat{P}_{jm}(\theta) = \sum_{i=1}^N w_i(\theta) Y_{jm}^{(i)}$$

donde las ponderaciones $w_i(\theta)$ son definidas para que sean máximas en $\theta = q_i$, y que decrezcan a cero cuando crece $|q_i - \theta|$. Los pesos $w_i(\theta)$ se definen a partir de los núcleos $K(u)$ como:

$$w_i(\theta) = \frac{K(\frac{\hat{\theta}_i - \theta}{h})}{\sum_k K(\frac{\hat{\theta}_k - \theta}{h})}$$

Luego,

$$(2.8) \quad \hat{P}_{jm}(\theta) = \frac{\sum_{i=1}^N K(\frac{\hat{\theta}_i - \theta}{h}) Y_{jm}^{(i)}}{\sum_{k=1}^N K(\frac{\hat{\theta}_k - \theta}{h})}$$

donde el denominador es la estimación de la densidad de las q_i . Esto implica que, como la distribución es conocida, se podría reemplazar el denominador por esta densidad conocida. Esto podría ocasionar valores de P levemente fuera del intervalo $[0,1]$, pero esto puede ser corregido.

Como la diferenciabilidad de \hat{P} depende de la diferenciabilidad de K , es ventajoso usar los núcleos cuadráticos o gaussianos, por ejemplo, si queremos estimar la función de información del test.

El parámetro h de ancho de banda debe ser elegido adecuadamente. Si h disminuye, también disminuye el sesgo, dado que sólo pocas observaciones muy cercanas a θ son efectivamente ponderadas, pero crece la varianza. Si h crece, disminuye la varianza, pues se ponderan más observaciones, pero aumenta el sesgo. En Hardle (1990) se dan varios procedimientos para elegir la ventana óptima a los efectos de minimizar el error cuadrático medio o funciones de este.

Douglas (1997) demuestra la consistencia conjunta de las CCI y de los parámetros del rasgo cuando el tamaño muestral y la longitud del test crecen juntos en un excelente artículo.

Considera un test de n ítems, administrado a N_n examinados, elegidos al azar. Se estiman las CCI de los n ítems y los N_n rasgos. Entonces ahora, se considera un nuevo test de $n + 1$ ítems, que no necesariamente contengan los n anteriores, y se administra a N_{n+1} examinados. La teoría asintótica estudia la adecuación de las curvas características del ítem y de los estimados del rasgo a medida que n tiende a infinito.

Cada test viene identificado por sus curvas características de los ítems $P_{n,i}$. Para cada curva $P_{n,i}$ se necesita construir un estimador $\hat{P}_{n,i}$ y similarmente construiremos estimadores $\hat{\theta}_{n,k}$ de los rasgos $\theta_{n,k}$. Notemos por \bar{P}_n a la curva característica del test, o sea, $\bar{P}_n(\theta) = \frac{1}{n} \sum_{i=1}^n P_{n,i}(\theta)$.

Hipótesis

A los efectos de obtener la consistencia de estas estimaciones, Douglas (1997) asume ciertas hipótesis sobre los ítems, las CCI, los tamaños muestrales N_n , los núcleos y los anchos de banda h_n .

1. Independencia local y unidimensionalidad.

2. Θ tiene distribución uniforme $(0,1)$.
3. \overline{P}'_n existe en $(0,1)$ y existe un número m tal que $0 < m < \overline{P}'_n(\theta)$ para todo $\theta \in (0,1)$.
4. Existe $C > 0$ tal que $np_{nk} < C$ para todo k donde $p_{nk} = P(\overline{U}_n = k)$.
5. Existen constantes M_L y M_U tal que $M_L n^{3/2} < N_n < M_U n^r$.
6. $P'_{n,i}$ existe y es continua en $(0,1)$ para todo n, i .
7. Para cada intervalo compacto $[a, b] \subset (0,1)$ existe una constante $M_{a,b}$ tal que para todo $\theta \in [a, b]$ n y todo n, i $|P'_{n,i}(\theta)| < M_{a,b}$.
8. K tiene soporte acotado, en particular existe una constante C_K , tal que $K(x) > 0$, cuando $|x| < C_K$ y $K(x) = 0$, cuando $|x| > C_K$.
9. K es Lipschitziana de constante L_K .
10. Existe $\alpha \in (0, 1/2)$ y números positivos L_α y U_α tal que para todo n , $L_\alpha n^{-\alpha} < h_n < U_\alpha n^{-\alpha}$.

Consistencia

A partir de estas hipótesis se obtiene primeramente:

Teorema 2.4.1. Dadas las hipótesis 1 a 7 $\max_{j=1,\dots,J_n} |\hat{\Theta}_{n,j} - \Theta_{n,j}| \rightarrow 0$ con probabilidad 1.

También se obtiene un resultado sobre el error cuadrático.

Teorema 2.4.2. Dadas las hipótesis 1 a 10 para $\theta_0 \in (0,1)$ fijo

$$(2.9) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(n^{-2\alpha})$$

donde α es la constante de la hipótesis 10.

Definiendo

$$\delta_{n,a,b} = \sup_{\theta \in (a,b)} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|$$

se obtiene:

Teorema 2.4.3. Dadas las hipótesis 1 a 10, $\delta_{n,a,b} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

Finalmente, prueba la consistencia conjunta pues a partir de los teoremas 3.6.1 y 3.6.3, y notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

Teorema 2.4.4. $\max\{\Delta_n, \delta_{n,a,b}\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

Para lograr la monotonía de los estimadores de esta sección hay una gran variedad de métodos, por ejemplo, el propuesto por Cheng y Lin (1981); Wright (1981); Friedman y Tibshirani (1984); Delecroix y Thomas-Agnan (2000); o Gijbels (2005) y Mammen (1991). Brunk (1955) propone un método para obtener estimadores con propiedades similares a los de la regresión no paramétrica. Ramsay (1998) y Kelly y Rice (1990) proponen usar splines, aunque este procedimiento es válido cuando la función de regresión satisface una ecuación de diferencias de segundo orden por lo que es, en general, no consistente. Hall y Huang (2001) usan un estimador del tipo núcleos con pesos modificados, Lee (2002) propone el uso de la regresión isótoma propuesta por Barlow, Bartholomew, Bremner y Brunk (1972); así como por Robertson, Wright y Dykstra (1988). Finalmente, Luzardo y Rodríguez (2015) proponen un método basado en estimar monótonamente la función inversa de las CCI.

2.5. Enfoque desde al análisis de datos funcionales

Ramsay (2016) presenta un muy interesante enfoque de los modelos de TRI no paramétrico a partir de variedades diferenciables. Pensar los modelos de TRI a partir del análisis de datos funcionales fue inicialmente introducido en Ramsay y Silverman (2005) y Ramsay et al. (2009). Dan herramientas para estimar curvas características de los ítems que sean suaves (diferenciables) y al mismo tiempo cumplan ciertos criterios sobre la forma. Ramsay y Silverman (2002) muestran este concepto prácticamente en el caso de un test.

Como ya hemos observado en la sección 1.3 la variable latente actúa como un índice y cada uno de sus valores tiene asociado un valor de probabilidad. Debido a la importancia de este argumento repetimos aquí el análisis de este concepto. Dado un valor de θ consideremos una función biyectiva h y entonces tenemos el transformado $\phi = h(\theta)$. Como h es biyectiva tiene inversa h^{-1} o sea $\theta = h^{-1}(h(\theta)) = h^{-1}(\phi)$

A partir de lo anterior tenemos que:

$$P(\theta) = P(h^{-1}(\phi)) = P^*(\phi)$$

donde $P^* = P \circ h^{-1}$, es decir, en el conjunto de las CCI se establece una relación de equivalencia mediante el hecho que dos CCI son equivalentes si son iguales ante transformaciones biyectivas del rasgo. Se pueden imponer sobre h restricciones adicionales como pedir que sea estrictamente creciente, o diferenciable. Entonces, para una presentación particular del modelo implica dos elecciones, el intervalo en el cual está definida θ y la medida de probabilidad (distribución) $dF(\theta)$ sobre ese intervalo. Es claro que esta elección influye en la forma de la curva.

Consideremos un test de n ítems dicotómicos. Se determina, entonces, un hipercubo cerrado en el espacio $[0, 1]^n$ al que Ramsay (1996) y Rosssi et al. (2002) llaman espacio de performance. La performance de los n ítems se modela mediante una curva \wp llamada variedad de performance. Cada punto de esta variedad se corresponde con cada nivel de performance, es decir $\wp = \wp(\theta) = (P_1(\theta), \dots, P_n(\theta))$. Si se tiene que cada ítem puede tener r_i respuestas se tiene que las diferentes respuestas son un punto de un simplejo de dimensión $r_i - 1$. La estructura del test entonces es un hipercubo simplicial de dimensión n donde cada borde es un simplejo. A partir de esto uno puede utilizar diferentes métricas para la variable latente dando curvas equivalentes. Por ejemplo, se puede utilizar una métrica probabilística (gaussiana, beta, etc), tomar como habilidad el rango del sujeto, la métrica de la longitud de arco y las sumas o la suma ponderada de los puntajes.

2.5.1. Estimacion invariante de la habilidad

Es claro que los estimadores de la habilidad deben ser invariantes respecto a transformaciones estrictamente monótonas. Una forma común es estimar θ mediante la suma de los puntajes en el test la cual cumple ese criterio. Esto se deduce del hecho que el valor esperado es la suma de las CCI de los ítems y estas son invariantes por lo visto anteriormente. La suma de puntajes de los ítems considera que todos ellos tienen la misma importancia en la medición del rasgo por lo que sería importante poder ponderar los ítems. Ramsay (2016) prueba que la estimación máxima verosimil del rasgo satisface este principio de invarianza.

La verosimilitud de las respuesta es:

$$(2.10) \quad P(U|\theta_j) = \prod_{i=1}^n P_i^{u_i} Q_i^{1-u_i} = \prod_{i=1}^n \left(\frac{P_i}{Q_i} \right)^{u_i} Q_i = \prod_{i=1}^n \left(\frac{P_i}{Q_i} \right)^{u_i} \left(1 + \frac{P_i}{Q_i} \right)^{-1}$$

donde $P_i = P_i(\theta)$.

Tomando logaritmo:

$$(2.11) \quad \ln L = \ln P(U_i|\theta_j) = \sum_{i=1}^n [u_i \ln \frac{P_i}{Q_i} - \ln(1 + e^{\ln \frac{P_i}{Q_i}})]$$

Si notamos por $W_i(\theta) = \ln \frac{P_i}{Q_i}$ se obtiene

$$(2.12) \quad \ln L(\theta) = \sum_{i=1}^n [u_i W_i(\theta) - \ln(1 + e^{W_i(\theta)})]$$

Derivando respecto a θ :

$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^n [u_i - P_i(\theta)] \frac{\partial W_i}{\partial \theta}$$

Como además si $\theta^* = h(\theta)$

$$\frac{\partial W_i}{\partial \theta^*} = \frac{\partial W_i}{\partial \theta} \frac{\partial h}{\partial \theta}$$

y a partir de la monotonía de h su derivada es distinta de 0 por lo que el principio de invarianza se cumple. El estimador obtenido tiene la ventaja que cada ítem está ponderado por las pendientes de las funciones W_i .

Esta ecuación sugiere usar el estimador ponderado

$$(2.13) \quad \hat{\theta} = n \left[\sum_{i=1}^n u_i \frac{\partial W_i}{\partial \theta} \right] / \left[\sum_{i=1}^n \frac{\partial W_i}{\partial \theta} \right]$$

Debido a la regla de la cadena podemos ver que el estimador es invariante con respecto a h .

Es claro que el valor esperado de $\hat{\theta}$ vive en el intervalo $[0, n]$

2.5.2. Métricas

La métrica de la longitud de arco viene dada por: $s(\theta) = \int_0^\theta \sqrt{\sum_i^n \left(\frac{dP_i}{dz}\right)^2} dz$ donde $z \in [0, n]$.

La distancia de la longitud de arco se puede interpretar a partir que la curva se define a partir de n coordenadas representando los distintos aspectos del test, una posición en la curva representa la performance total y la logitud de curva desde este punto hasta su término representa la mejora necesaria para adquirir la maestría.

Para definir una métrica en el intervalo $[0, n]$ Ramsay y Wiberg (2016) proponen la distribución $TS\beta$ que tiene densidad

$$p(S|\alpha, \beta, h_0, h_n) = \frac{h_0(1-S/n)+h_1(S/n)+(S/n)^{\alpha-1}(1-S/n)^{\beta-1}}{n(h_0+h_n)/2+B(\alpha, \beta)}$$

donde $B(\alpha, \beta)$ es la función beta.

2.5.3. Algoritmo para el análisis funcional de un test

Ramsay (2016) propone el siguiente algoritmo para analizar un test usando una aproximación al análisis de datos funcionales.

Paso 1 Inicialización

a) *Estimación inicial de θ_p* : Recomienda usar la suma total de los ítems en el test, aunque se puede usar otra.

b) *Inicializar la densidad*: Inicializar la densidad $TS\beta$ sobre el intervalo $[0, n]$.

Paso 2 Iteración Para cada iteración $v = 1, 2, \dots$ se hacen los siguientes pasos:

- a) *Convertir rangos en quantiles* Los rangos normalizados $\tilde{u} = p/(N + 1)$ se transforman en $\hat{\theta}_p^{v-1}$ usando la función de quantiles $Q^{v-1}(u)$ de la distribución estimada en la iteración previa.
- b) *Calcular los extremos de los intervalos* Sea K el número de intervalos que se desea. Si el dominio es cerrado se calculan los valores $Q^{v-1}((0 : 2k)/2K)$ y en otro caso se calculan $Q^{v-1}((0 : 2k)/(2K - 1))$ más puntos extremos adecuados. Las cuantiles impares serán los centros de los intervalos y las pares los extremos de los intervalos.
- c) *Cálculo de W_i y $\hat{\theta}$* : Se rompen los empates de los estimadores de la habilidad si existen agregando un ruido aleatorio. Se ordenan y les asigna a cada uno su cuantila p correspondiente. Los promedios dentro de cada intervalo se usan como estimaciones de las probabilidades y a partir de ellos se estiman los $W_i(\theta)$ en los centros de los intervalos.
- d) *Estimación mediante splines del logaritmo del odds*: Se usa una base de splines de orden 4 con nudos en los centros de los intervalos para realizar un suavizamiento con splines de los valores hallados en 2c.
- e) *Actualización de la nueva habilidad*: Se actualizan los valores de $\hat{\theta}_p^{v-1}$ aplicando la estimación ponderada de la habilidad usando los valores de W_i del paso 2c.
- f) *Actualización de la densidad*.

Capítulo 3

Modelo no paramétrico multidimensional

3.1. Introducción

En este capítulo presentaremos dos modelos no paramétricos multidimensionales. Para esto se utilizarán estimadores de regresión no paramétrica con variables independientes multidimensionales. Existen varias formas de encarar el problema de la regresión no paramétrica, los más populares son aquellos basados en núcleos, los basados en splines y los que utilizan wavelets. Cada uno de estos métodos tienen sus propias fortalezas y debilidades pero los que están basados en núcleos se caracterizan por su simplicidad matemática, además de tomar un formato intuitivo, como ser promedios ponderados locales.

En principio es natural pensar en aplicar un estimador semejante al que utilizó Ramsay para el caso unidimensional para ítems multidimensionales, con una dificultad adicional al tener que estimar un rasgo que ahora es un vector.

Consideremos entonces variables de respuestas dicotómicas que dependen de un rasgo multidimensional Θ . Es claro que las respuestas a un ítem condicionadas a un valor del rasgo son variables aleatorias Bernoulli, por lo que la probabilidad $p(\theta)$ es la esperanza condicional de la respuesta. Esta esperanza de la respuesta condicionada a θ es lo que conocemos por regresión. Por lo tanto es razonable pensar en estimar la CCI mediante la regresión no paramétrica. Hardle (1990) menciona que la estimación no paramétrica de las curvas de regresión tiene cuatro objetivos fundamentales: “Primero provee un método versátil para explorar la relación entre dos variables. Segundo, da predicciones de las observaciones sin hacer referencia a un modelo paramétrico fijo. Tercero, provee una herramienta para encontrar observaciones espúreas estudiando la influencia de puntos aislados. Cuarto, constituye un método flexible para sustituir valores perdidos o interpolar entre valores de X ”.

La idea general de la regresión no paramétrica es la del promedio local, es decir, estimar la esperanza condicional mediante un promedio ponderado, donde los pesos son una función que depende del punto donde se va a estimar y de los valores observados.

En este capítulo se presentará una familia de estimadores mediante polinomios locales con núcleos. Esta familia incluye al estimador de Ramsay (Ramsay 1991) en el caso unidimensional y su generalización multidimensional (Luzardo y Forteza 2014), y adicionalmente genera una gran variedad de posibilidades. El modelo de Ramsay y su generalización corresponderán al modelo polinomial de grado 0. Observaremos con atención al polinomio de grado 1 ya que en el caso de variables observables tiene mejores propiedades asintóticas y de frontera que los estimadores de

Nadaraya-Watson.

Lo primero a notar es que los rasgos se ignoran por lo tanto deben estimarse. Para ello se recurre a un procedimiento no paramétrico valiéndose de la función de distribución empírica de variables auxiliares. Se necesitan tantos estadísticos para ordenar a los sujetos como rasgos y esto presenta complicaciones técnicas que detallaremos en el capítulo de consistencia. En el caso unidimensional se puede utilizar el promedio o el puntaje total, pero en el caso multidimensional no es inmediato determinar cómo deben ser los estadísticos y que condiciones deben cumplir para que la estimación sea consistente.

Comenzaremos presentando el modelo indicando como se estimarán los rasgos y como se estimarán las curvas características del ítem y luego obtendremos tres tipos de resultados estadísticos para este modelo.

Un primer tipo de resultados es obtener estimadores para la distribución del rasgo concretamente de la densidad conjunta y la cópula subyacente. Creemos que este punto es muy interesante pues el método solo va a presuponer el conocimiento de las marginales (ni siquiera esto es necesario) y a partir de esto se puede reconstruir la distribución conjunta. Para lograr la demostración de la consistencia del estimador de la densidad y su distribución asintótica deberemos suponer que el rasgo es medido sin error aunque seguiremos suponiendo que los sujetos son elegidos aleatoriamente de la población. En el capítulo de simulaciones veremos que el estimador de la densidad conjunta funciona muy bien aun cuando el rasgo sea medido con error.

También presentamos la consistencia conjunta de la estimación del rasgo y de la CCI en el caso multidimensional encontrando las condiciones que deben cumplirse para que esta ocurra. También encontraremos una acotación exponencial para el error cuadrático medio. Esta sección la cual consideramos central, presenta dificultades técnicas importantes, debiéndose usar fuertemente las desigualdades de concentración de desarrollo reciente. Mencionamos que como producto adicional al investigar las condiciones antes dichas se logró definir una familia de funciones a las que llamaremos asintóticamente Liptchiz.

Finalmente para el caso de la habilidad medida sin error se deducirá la distribución asintótica del estimador de las CCI así como la consistencia completa, lo cual constituirá el tercer bloque de resultados.

3.2. Presentación del modelo multidimensional

Comencemos viendo la estructura general del modelo y como se estimará el rasgo, luego se mostrará como se estimaran las CCI, lo cual se basa en la regresión multivariada usando núcleos.

A los efectos de estimar cada componente del vector que representa un rasgo específico, utilizaremos estadísticos para ordenar a los sujetos y así tomar como estimación de dicha componente del rasgo en cada sujeto el cuartil empírico correspondiente. Expongamos primeramente el método y luego lo formularemos en forma precisa. Se definirá para cada rasgo un estadístico (al que notaremos $g_{n,l}$). Estos estadísticos dependerán solo de un rasgo, al cual estimarán, o sea son independientes de los restantes rasgos. En cada componente se ordenarán los puntajes obtenidos en el estadístico $g_{n,l}$ y luego se tomará el cuartil empírico para obtener el rasgo estimado del sujeto.

Formalmente consideramos ahora que el rasgo es un vector aleatorio en R^d

$$\Theta = (\Theta_1, \dots, \Theta_d).$$

Sea (Ω, A, P) un espacio de probabilidad y consideremos sin pérdida de generalidad que

el rasgo latente Θ tiene distribuciones marginales uniformes $U[0, 1]$. Es claro que todo lo que deduzcamos para este supuesto funcionará para cualquier tipo de distribución marginal.

Es interesante independizarse de la posible distribución del rasgo, ya que esta, usualmente, es desconocida.

Comencemos viendo esto en el caso unidimensional, para ello notemos por un momento al rasgo como τ y supongamos que tiene distribución F . Sabemos que $F(\tau)$ se distribuye uniforme $[0, 1]$. Por lo tanto, podríamos estimar en vez de τ una transformación monótona de ella $\theta = F(\tau)$.

Así pues, se puede suponer que el rasgo tiene distribución uniforme, y entonces, este rasgo se estima mediante la proporción empírica.

Si conocemos la distribución F del rasgo, obtenemos la escala original, mediante $\tau = F^{-1}(\theta)$, que tiene distribución F . En el caso multidimensional recurrimos a las cópulas.

Sea $(\Theta_1, \dots, \Theta_d)$ y supongamos conocidas sus marginales y que son continuas. El vector $(U_1, U_2, \dots, U_d) = (F_1(\Theta_1), \dots, F_d(\Theta_d))$ tiene marginales uniforme. La cópula de $(\Theta_1, \dots, \Theta_d)$ se define como

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d] = P(\Theta_1 \leq F_1^{-1}(u_1), \dots, \Theta_d \leq F_d^{-1}(u_d))$$

Esta función contiene toda la información de dependencia del vector de rasgos. Aplicando el teorema de Sklar podemos reconstruir a partir de las marginales la función de distribución de los rasgos.

Si notamos por $H(\theta_1, \dots, \theta_d) = P[\Theta_1 \leq \theta_1, \dots, \Theta_d \leq \theta_d]$ la distribución conjunta del rasgo entonces $H(\theta_1, \dots, \theta_d) = C(F_1(\theta_1), \dots, F_d(\theta_d))$. Por lo que de aquí en más trabajaremos con marginales uniformes.

Consideremos una sucesión de funciones vectoriales, Borel medibles, \mathbf{g}_n en R^n a valores en $[0, 1]^d$. Sea $g_{n,l}$ la componente l -ésima de \mathbf{g}_n , para $x \in R^n$ $\mathbf{g}_n(x) = (g_{n,1}(x), \dots, g_{n,d}(x))$.

Consideremos un test de n ítems contestados por N sujetos y las variables aleatorias $X_{i,k}$, $i = 1, \dots, n$, $k = 1, \dots, N$ que indican la respuesta del k -ésimo sujeto al i -ésimo ítem. Estas variables dependen de d rasgos latentes (Θ_l) y sea $\mathbf{X} = (X_1, \dots, X_n)$ el vector aleatorio que indica las respuestas a los n ítems. Específicamente, \mathbf{X}_k indicará el vector de respuestas del sujeto k .

Supongamos que las funciones $g_{n,l}(X)$ son independientes de $\Theta_1, \dots, \Theta_{l-1}, \Theta_{l+1}, \dots, \Theta_n$ para todo $1 \leq l \leq n$. Definamos la sucesión de funciones $\{G_{g_n}\} : R^d \rightarrow R^d$ tal que para cada $x = (x_1, \dots, x_d)$ se tiene:

$$(3.1) \quad G_{g_n}(x) = (P(g_{n,1}(X) \leq x_1), \dots, P(g_{n,d}(X) \leq x_d))$$

Sean las distribuciones empíricas: $\hat{F}_{N,l}(x) = \frac{\#\{g_{n,l}(X) \leq x\}}{N} = \frac{\sum_{i=1}^N \chi_{\{g_{n,l}(X_i) \leq x\}}}{N}$ con $l = 1, \dots, d$ y definamos la función:

$$(3.2) \quad \hat{G}_N(x_1, \dots, x_d) = (\hat{F}_{N,1}(x_1), \dots, \hat{F}_{N,d}(x_d))$$

A partir de lo anterior, para estimar la componente l -ésima del rasgo se utiliza la función $g_{n,l}$ para ordenar los sujetos. Luego tomamos la función de distribución empírica, es decir,

$$(3.3) \quad \hat{\Theta}_n = \hat{G}_N(g_n(X))$$

Observemos que no depende este método de estimación del tipo de respuesta obtenido, pudiendo estas ser dicotómicas, politómicas o continuas. Si conocemos las marginales, podemos tener estimaciones en la escala original aplicando la inversa de cada una de estas distribuciones al rasgo estimado mediante la distribución uniforme.

3.2.1. Estimación de la CCI

Dado que $P(\theta_i) = E(X|\Theta = \theta_i)$ podemos considerar que el modelo subyacente es:

$$P(\theta_i) = m(\theta_i) + \sigma(\theta_i)\varepsilon_i \text{ con } i = 1, \dots, N$$

y donde los ε_i condicionados a $\Theta_1, \dots, \Theta_N$ son variables aleatorias independientes con media 0 y varianza 1. Supongamos que la densidad d-dimensional de Θ es $f(\theta)$ y que tenemos N sujetos con rasgos $\theta_1, \dots, \theta_N$. Sea H una matriz de dimension d, simétrica y no singular.

Se puede estimar esta densidad no paramétricamente mediante:

$$\hat{f}_{n,H}(\theta) = \frac{1}{N|H|} \sum_{i=1}^N K(H^{-1}(\theta_i - \theta)) = \frac{1}{N} \sum_{i=1}^N K_H(\theta_i - \theta)$$

donde $|H|$ es el determinante de H y $K_H(\theta_i - \theta) = \frac{1}{|H|} K(H^{-1}(\theta_i - \theta))$.

La matriz H contiene los parámetros de ancho de la ventana, es decir, controla cuán rápidamente los pesos decrecen a cero. Si queremos utilizar un solo ancho de ventana podemos tomar $H = hI_d$ donde I es la matriz identidad $d \times d$ en tanto si se quiere tener una ventana diferente para cada dimensión se puede tomar $H = \text{diag}(h_1, \dots, h_d)$.

Quedan en estos casos

$$\hat{f}_{n,h}(\theta) = \frac{1}{Nh^d} \sum_{i=1}^N K(h^{-1}(\theta_i - \theta))$$

$$\hat{f}_n(\theta) = \frac{1}{Nh_1 h_2 \dots h_d} \sum_{i=1}^N K(h_1^{-1}(\theta_{i1} - \theta_1), \dots, h_d^{-1}(\theta_{id} - \theta_d))$$

Supongamos que tenemos N respuestas a un ítem (cuya respuesta sin pérdida de generalidad supondremos continua) $(\theta_1, X_1), \dots, (\theta_N, X_N)$, que son independientes, idénticamente distribuidas con densidad conjunta $f_{\Theta,X}(\theta, x)$.

La densidad condicional de X dada Θ es: $f_{X|\Theta} = \frac{f_{\Theta,X}(\theta, x)}{f_{\Theta}(\theta)} = \frac{f_{\Theta,X}(\theta, x)}{\int f_{\Theta,X}(\theta, x) dx}$

Así tenemos que:

$$(3.4) \quad E(X | \Theta = \theta) = \frac{\int x f_{\Theta,X}(\theta, x) dx}{\int f_{\Theta,X}(\theta, x) dx}$$

Por lo tanto, para estimar la esperanza condicional utilizamos:

$$(3.5) \quad E(\widehat{X | \Theta} = \theta) = \frac{\int x \widehat{f_{\Theta,X}}(\theta, x) dx}{\int \widehat{f_{\Theta,X}}(\theta, x) dx}$$

Consideremos ahora un núcleo R^{d+1} para estimar la densidad condicional definido como $G_h(\cdot)K_H(\cdot)$ donde G es un núcleo unidimensional y K un núcleo de dimensión d.

El estimador por núcleos de la densidad conjunta es:

$$\widehat{f_{\Theta,X}}(\theta, x) = \frac{1}{N} \sum_{i=1}^N G_h(X_i - x) K_H(\theta_i - \theta)$$

realizando el cambio de variable, $-u = \frac{X_i - x}{h}$

$$N^{-1} \sum_{i=1}^N \int x \frac{1}{h} G\left(\frac{X_i - x}{h}\right) K_H(\theta_i - \theta) dx = N^{-1} \sum_{i=1}^N X_i K_H(\theta_i - \theta)$$

Además,

$$\int \widehat{f_{\Theta,X}}(\theta, x) dx = N^{-1} \sum_{i=1}^N K_H(\theta_i - \theta)$$

Obtenemos finalmente como estimador:

$$(3.6) \quad E(\widehat{X | \Theta} = \theta) = \frac{N^{-1} \sum_{i=1}^N X_i K_H(\theta_i - \theta)}{N^{-1} \sum_{i=1}^N K_H(\theta_i - \theta)} = \frac{\sum_{i=1}^N X_i K_H(\theta_i - \theta)}{\sum_{i=1}^N K_H(\theta_i - \theta)}$$

Notemos que el numerador es la convolución de K con los datos y el denominador es la convolución de K con la unidad. Se puede mostrar que si las respuestas son dicotómicas el estimador se mantiene.

Este es un estimador del tipo de promedio local ponderado con pesos, $w_i(x) = \frac{K_H(\theta_i - \theta)}{\sum_{i=1}^N K_H(\theta_i - \theta)}$. Es decir, es un promedio ponderado de las respuestas de aquellos cuya θ_i vive en una bola o cubo alrededor de θ .

Es claro que en el caso general, se ignoran los valores de $\theta_1 \dots \theta_N$. Pero si se tiene una sucesión de estimadores $\hat{\theta}_{n,i}$ que converge a θ_i para todo $i = 1, \dots, N$ en probabilidad, casi seguramente o completamente es inmediato pensar en sustituir θ_i por $\hat{\theta}_i$. Probaremos luego que esto se puede hacer con los estimadores del rasgo definidos en la sección anterior.

Por lo tanto, un estimador de la CCI del ítem j es:

$$(3.7) \quad \widehat{P_j}(\theta) = \frac{\sum_{i=1}^N X_{i,j} K_H(\hat{\theta}_i - \theta)}{\sum_{i=1}^N K_H(\hat{\theta}_i - \theta)}$$

Es inmediato verificar que el estimador encontrado es la solución de

$$(3.8) \quad \min_{\beta_0} \sum_{i=1}^N (X_{i,j} - \beta_0)^2 K_H(\hat{\theta}_i - \theta)$$

O sea, es un estimador localmente constante. Reemplazando β_0 por un polinomio podemos encontrar estimadores de regresión basados en núcleos del tipo polinomios locales. Un caso importante es el polinomio de grado 1 el cual es la solución de:

$$(3.9) \quad \min_{\beta_0, \beta_1} \sum_{i=1}^N (X_{i,j} - \beta_0 - (\hat{\theta}_i - \theta)^t \beta_1)^2 K_H(\hat{\theta}_i - \theta)$$

Usando mínimos cuadrados ponderados la solución queda:

$$(3.10) \quad \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X \text{ donde}$$

$$\Sigma = \begin{pmatrix} 1 & (\hat{\theta}_1 - \theta)^t \\ \vdots & \vdots \\ 1 & (\hat{\theta}_N - \theta)^t \end{pmatrix}, X = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{N,j} \end{pmatrix}$$

$$W = \begin{pmatrix} K_H(\hat{\theta}_1 - \theta) & 0 & \dots & 0 \\ 0 & K_H(\hat{\theta}_2 - \theta) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & K_H(\hat{\theta}_N - \theta) \end{pmatrix}$$

En este caso β_0 es el estimador de la CCI y β_1 las derivadas parciales con respecto a las componentes de θ . Al estimador de la CCI obtenido mediante polinomios de grado 1 lo notaremos como

$$\widehat{P_j^1}(\theta) = \hat{\beta}_0$$

Utilizando el vector $e_1 = (1, 0, 0, \dots, 0)$ de la base canónica entonces

$$(3.11) \quad \widehat{P_j^1}(\theta) = e_1^t (\hat{\beta}_0, \hat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X$$

Hemos entonces obtenido dos estimadores de las CCI dados por las ecuaciones 3.7 y 3.11

3.2.2. Respuestas politómicas y continuas

Si las respuestas son politómicas podemos aplicar el método generando tantas variables dicotómicas como opciones tiene el ítem. Es decir X_{ij}^m valdrá 1 si el sujeto i en el ítem j eligió la opción m . Luego se aplica el estimador.

Si tenemos un modelo de respuesta graduada podemos considerar variables aleatorias dicotómicas como: $Y_{ij}^m = \chi_{(X_{ij} \leq m)}$, luego aplicar el modelo y calcular:

$$(3.12) \quad P(X = m | (\theta_1, \dots, \theta_d)) = P(X \leq m) - P(X \leq (m - 1))$$

Si hay respuestas continuas entonces podemos definir para un $x \in R$ variables $Y_{ij} = \chi_{(X_{ij} \leq x)}$. Con estas variables podemos estimar la distribución condicionada al rasgo de las respuestas.

También en el caso continuo se puede estimar la esperanza condicional de la respuesta, usando el estimador con la respuesta original (es decir no convirtiéndola en una indicatriz para cada x).

3.3. Estimación de la densidad conjunta del rasgo

Surge como primer objetivo estimar la densidad conjunta de las habilidades a partir de las habilidades estimadas. Es claro que el problema que se presenta pretende contestar lo siguiente: Si $\Theta = \hat{\Theta} + \epsilon$ como hallar la densidad de Θ a partir de la densidad de $\hat{\Theta}$. Hay trabajos en este sentido, en especial un artículo de Sperlich (2009). Sin embargo presupone conocimiento sobre el error de estimación elemento del que carecemos. En el apartado de limitaciones trataremos nuevamente este tema. Ahora lo abordaremos desde la perspectiva simplificada que la habilidad es estimada sin error, o que si estimamos la densidad de $\hat{\Theta}$ esta se encontrará cercana a la densidad de Θ . Como en este capítulo demostraremos la consistencia de la estimación del rasgo esta es una aproximación aceptable, aunque es claro que no estamos tomando en cuenta un aumento del sesgo y la varianza del error.

La estructura puede ser estimada tanto con los rasgos estimados en el cubo $[0, 1]^d$ o ya transformados mediante la inversa de las marginales. En lo que sigue supongamos que F_i , $i=1, \dots, d$ son las marginales y que Θ_{ij} $i=1, \dots, d$ $j=1, \dots, N$ ha sido obtenido aplicando F_i^{-1} al estimado del rasgo a partir del cuartil empírico.

Es claro que solo tiene sentido práctico aplicar la estimación no paramétrica de la densidad del rasgo en el caso de dimensión mayor o igual a dos, pues es en este caso que solo suponemos el conocer las marginales (aunque solo se supongan uniformes) en tanto en el caso unidimensional se supone conocida la distribución. Igualmente comenzaremos presentando el estimador unidimensional por la simplicidad de los cálculos para luego generalizar.

La gráfica que sigue muestra la proximidad de la densidad real y la estimada usando los rasgos estimados, esto nos indica que la aproximación usada es adecuada.

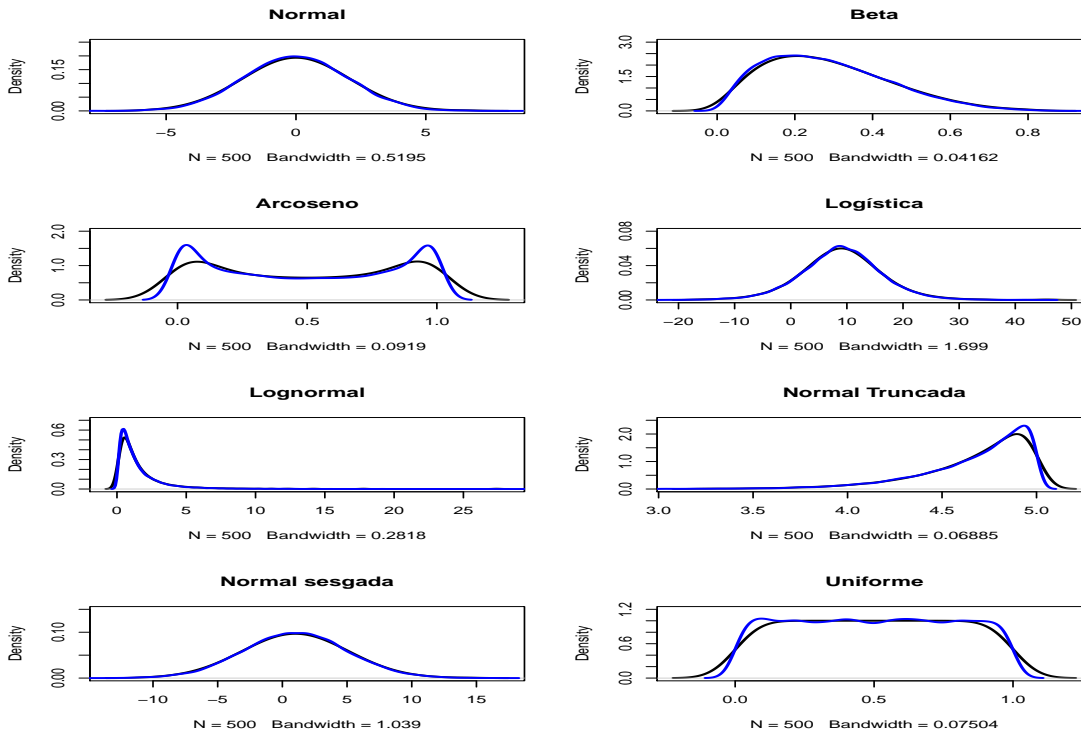


Figura 3.1: Proximidad de la densidad real y estimada por tipo de distribución

Caso unidimensional

Vamos a usar el estimador usual de la densidad por núcleos pero reemplazando la variable independiente por el estimador obtenido recién.

Supongamos que nuestros N sujetos constituyen una muestra del rasgo, o sea $\Theta_1, \dots, \Theta_N$ son variables aleatorias independientes idénticamente distribuidas y estas variables tienen una densidad común $f(\theta)$. En este caso consideramos que Θ está medido sin error por $\hat{\Theta}$

Nuestro objetivo es, a partir de esta muestra, estimar $f(\theta)$. Para ello consideremos la familia \mathfrak{F} de las funciones de distribución absolutamente continuas con densidad f la cual queremos encontrar.

Recordemos que $\hat{f}_{N,h}(\theta)$ se reduce a una convolución $\hat{f}_{N,h} = \bar{K}_h * F_N$ donde $\bar{K}_h(z)$ es un núcleo. El estimador toma la forma

$$(3.13) \quad \hat{f}_{N,h}(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\theta}_i}{h}\right)$$

Como el estimador de la densidad es la convolución de un núcleo con la distribución empírica, para demostrar la consistencia del estimador es clave obtener resultados sobre la convolución de un núcleo con una función. El siguiente lema nos da condiciones para que la convolución converja puntualmente cuando la ventana tiende a 0. Este resultado será luego usado en el teorema 3.3.2

Lema 3.3.1. Sea K una función real que cumple:

- i) K es acotada y simétrica
- ii) $\int |K(t)| dt = C_1 < \infty$
- iii) $\lim_{\theta \rightarrow \pm\infty} |\theta K(\theta)| = 0$

Sea g una función real tal que $\int |g(t)| dt = C_2 < \infty$ y consideremos la convolución:

$$g_h(\theta) = K_h * g(\theta) = \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) g(t) dt \text{ donde } K_h(z) = \frac{1}{h} K(z/h) \text{ entonces}$$

$$g_h(\theta) \longrightarrow g(\theta) \int K(t) dt \text{ si } h \rightarrow 0 \text{ para todo punto } \theta \text{ de continuidad de } g$$

Demostración: Sea θ un punto de continuidad de g . Dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$|g(u) - g(\theta)| < \frac{\epsilon}{3C_1} \text{ si } |\theta - u| < \delta.$$

Como K es simétrico: $|g_h(\theta) - g(\theta) \int K(t) dt| = \left| \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) g(t) dt - g(\theta) \int K(t) dt \right| =$
 $\left| \int K(-u) g(\theta + uh) du - g(\theta) \int K(u) du \right| \leq \int |K(u)| |g(\theta + uh) - g(\theta)| du$. Ahora

$$\int |K(u)| |g(\theta + uh) - g(\theta)| du < \frac{\epsilon}{3C_1} \int_{|u| < \delta/h} |K(u)| du +$$

$$\int_{|u| \geq \delta/h} |K(u)| |g(\theta)| du + \int_{|u| \geq \delta/h} |K(u)| |g(\theta + uh)| du$$

Sabemos que θ es un punto de continuidad de g , entonces $|g(\theta)| < C_3$, además

$$\int_{|u| \geq \delta/h} |K(u)| du < \frac{\epsilon}{3C_3} \text{ pues es la cola de una integral convergente. Por último,}$$

$$\int_{|u| \geq \delta/h} |K(u)| |g(\theta + uh)| du \leq \frac{h}{\delta} \int_{|u| \geq \delta/h} |uK(u)| |g(\theta + uh)| du \leq$$

$$\frac{h}{\delta} \frac{\epsilon \delta}{3C_2} \int_{|u| \geq \delta/h} |g(\theta + uh)| du \leq \frac{\epsilon}{3C_2} \int |g(t)| dt = \frac{\epsilon}{3} \text{ pues } |uK(u)| < \frac{\epsilon \delta}{3C_2}$$

para h suficientemente chico.

A partir de estas tres desigualdades obtenemos que $|g_h(\theta) - g(\theta) \int K(t) dt| < \epsilon$ para $h \rightarrow 0$

Veremos ahora un teorema que nos garantiza la consistencia del estimador de la densidad.

Para ello se probará que $E(f_N(\theta)) \xrightarrow{h} f(\theta)$ y $Var(f_N(\theta)) \rightarrow 0$ lo cual garantiza la consistencia

o sea que $f_N(\theta) \xrightarrow{P} f(\theta)$.

Teorema 3.3.2. Supongamos que el rasgo es medido sin error o sea $\widehat{\Theta}_i = \Theta_i$ para todo i . Sea θ un punto de continuidad de f , K un núcleo simétrico y acotado y tal que $\int K(t) dt = 1$. Además $|\theta K(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$. Entonces

$$f_N(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\Theta}_i}{h}\right) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right)$$

es un estimador asintóticamente insesgado de $f(\theta)$ si $h \rightarrow 0$. Si además $Nh \rightarrow \infty$ es estimador consistente de $f(\theta)$.

Demostración: Por el lema 3.3.1 $E(f_N(\theta)) = \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) f(t) dt \xrightarrow{h} f(\theta)$

$$Var(f_N(\theta)) = \frac{1}{Nh^2} Var(K\left(\frac{\theta - X_1}{h}\right)) \leq \frac{1}{Nh^2} E(K^2\left(\frac{\theta - X_1}{h}\right)) = \frac{1}{Nh} \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow \frac{f(\theta)}{Nh} \int K^2(t) dt$$

Esta convergencia es válida ya que se cumple $|\theta K(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, entonces, $|\theta^2 K^2(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, lo que implica que $|\theta K^2(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, pues $|\theta K^2(\theta)| \leq |\theta^2 K^2(\theta)|$ para $|\theta| > 1$. Además $\int |K^2(t)| dt \leq \|K\|_\infty \int |K(t)| dt < \infty$. Entonces

si $Nh \rightarrow \infty$ $Var(f_N(\theta)) \rightarrow 0$. Hemos demostrado que si $Nh \rightarrow \infty$ y $h \rightarrow 0$, entonces,

$$f_N(\theta) \xrightarrow{P} f(\theta)$$

Observación: En realidad hemos probado la convergencia en L^2 , es decir que $E(f_N(\theta) - f(\theta))^2 \rightarrow 0$

Observación: el orden de la varianza es $\frac{1}{Nh}$, pues,

$$Var(\sqrt{Nh}f_N(\theta)) = NhVar(f_N(\theta)) = \frac{1}{h}Var(K(\frac{\theta-X_1}{h})) = \frac{1}{h}E(K^2(\frac{\theta-X_1}{h})) - \frac{1}{h}E^2(K(\frac{\theta-X_1}{h}))$$

luego,

$$Var(\sqrt{Nh}f_N(\theta)) \rightarrow f(\theta) \int K^2(u)du \text{ si } h \rightarrow 0 \text{ ya que}$$

$$\frac{1}{h^2}E^2(K(\frac{\theta-X_1}{h})) \rightarrow (f(\theta) \int K(u)du)^2 \text{ y } \frac{1}{h}E(K^2(\frac{\theta-X_1}{h})) \rightarrow f(\theta) \int K^2(u)du$$

de donde $\frac{Var(f_N(\theta))}{1/Nh} \rightarrow C$

Las proposiciones 3.3.3 y 3.3.4 que siguen dan aproximaciones al error cuadrático medio (ECM) calculando el sesgo y la varianza y a partir de esto se obtiene la ventana óptima en cada caso. La proposición 3.3.3 es más exigente que la 3.3.4.

Proposición 3.3.3. Sea θ un punto de continuidad de f y que K cumple las hipótesis del lema 3.3.1. Además suponemos que:

- i) existe f'' y $\sup|f''| < M$ y que el núcleo cumple:
- ii) $\int uK(u)du = 0$ y
- iii) $\int u^2K(u)du = a < \infty$

Entonces la ventana óptima cumple: $h \cong C_\theta N^{-1/5}$ y $ECM \cong CN^{4/5}$

Demostración: Sabemos que $ECM = B^2 + Var$ donde B indica el sesgo. El sesgo es:

$$B(f_N(\theta)) = E(f_N(\theta)) - f(\theta) = \frac{1}{h} \int K(\frac{\theta-t}{h})f(t)dt - f(\theta) =$$

$$\int K(u)(f(\theta + hu) - f(\theta))du = \int K(u) \left[f'(\theta)hu + f''(\xi)\frac{h^2u^2}{2} \right] du = \frac{h^2}{2} \int u^2K(u)f''(\xi)du \leq \frac{Mah^2}{2}$$

Si tomamos $f(\theta) = \theta^2$ entonces la desigualdad se transforma en igualdad, y $B_N(f_N(\theta)) \cong Ch^2$

Tenemos pues que $ECM = \frac{C_1}{Nh} + C_2h^4 + o(\frac{1}{Nh} + h^4)$ entonces el h óptimo sería $h \cong CN^{-1/5}$ y

$$ECM \cong CN^{-4/5}$$

.

Proposición 3.3.4. Consideremos que θ es un punto de continuidad de f y K cumple las hipótesis de los lemas anteriores. Además

- i) f es Lipchitz de constante L , es decir pido menos regularidad que en los lemas anteriores.
- ii) $\int |K(u)u|du = a_3 < \infty$

Entonces la ventana óptima será $h = CN^{-1/2}$ y el $ECM(f_N(\theta)) \cong CN^{-2/3}$

Demostración: A partir de las hipótesis ya sabemos que:

$$Var(\sqrt{Nh}f_N(\theta)) \rightarrow f(\theta) \int K^2(t)dt$$

$$\begin{aligned} \text{Ya que } \int K = 1 \text{ el sesgo es: } |B(f_N(\theta))| &= |E(f_N(\theta)) - f(\theta)| = \left| \frac{1}{h} \int K(\frac{\theta-\mu}{h})f(u)du - f(\theta) \right| \leq \\ &\leq \int |K(t)||f(\theta + th) - f(\theta)|dt \leq L \int |th||K(t)|dt = Lh \int |tK(t)|dt = La_3h \end{aligned}$$

así pues $|B(f_N(\theta))| \leq La_3h$

Si utilizo una función lineal $f(t) = LtI_{[\theta-c, \theta+c]}(t)$ obtengo la igualdad, entonces:

$$B(f_N(\theta)) \cong hC$$

luego $ECM = \frac{C_1}{Nh} + C_2h^2 + o(\frac{1}{Nh} + h^2)$ entonces el orden óptimo es: $\frac{1}{Nh} \cong h^2$ de donde $h = CN^{-1/3}$ así pues

$$ECM(f_N(\theta)) \cong CN^{-2/3}$$

Observación: A menor regularidad se obtiene menor velocidad de convergencia. Esto no tiene nada que ver con la parte aleatoria del problema.

Observemos que al sesgo le conviene que la ventana sea muy chica en tanto a la varianza que sea grande, es decir, existe un compromiso permanente entre sesgo y varianza.

3.3.1. Caso del rasgo multidimensional

La forma general del estimador de la densidad del rasgo multidimensional se obtiene como :

$$(3.14) \quad \hat{f}_{N,H}(\theta) = \frac{1}{N} \sum_{i=1}^N K_H(\theta - \hat{\Theta}_i) \text{ donde } K_H(\theta) = |H|^{-1} K(H^{-1}\theta)$$

y H es una matriz $d \times d$ simétrica definida positiva llamada matriz de ancho de ventanas o de banda y K es un núcleo de dimension d . La matriz H tiene $d(d+1)/2$ coeficientes independientes lo cual para dimensiones altas incrementa sustancialmente el número de parámetros. Hay dos casos particulares que se obtienen tomando $H = \text{diag}(h_1, \dots, h_d)$ y $H = hI$ con I la matriz identidad $d \times d$. En el primer caso se obtiene el estimador

$$\hat{f}_{N,H}(\theta) = \frac{1}{N \prod_{j=1}^d h_j} \sum_{i=1}^N K\left(\frac{\theta_1 - \Theta_{i1}}{h_1}, \dots, \frac{\theta_d - \Theta_{id}}{h_d}\right)$$

En el segundo caso queda:

$$\hat{f}_{N,H}(\theta) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{\theta_1 - \Theta_{i1}}{h}, \dots, \frac{\theta_d - \Theta_{id}}{h}\right) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right)$$

Veamos ahora una versión multidimensional del lema de Bochner. Se puede usar este resultado para obtener los teoremas vistos para cuando el rasgo es unidimensional en el caso multidimensional más sencillo. Este lema es el análogo al lema 3.3.1 para el caso multidimensional.

Lema 3.3.5. Sea $K_h = \frac{1}{h^d} K(\frac{\bullet}{h})$. Consideremos f en $L^1(R^d)$ y K en $L^1(R^d) \cap L^\infty(R^d)$. Además:

i) $\int_{R^d} K(t) dt = 1$

ii) $K(\theta) = o(\|\theta\|^{-d})$ si $\|\theta\| \rightarrow +\infty$

Sea $f_h = f * K_h$ entonces $f_h \rightarrow f$ si $h \rightarrow 0$ en cada punto de continuidad θ de f .

Demostración: Al ser θ un punto de continuidad de f , dado $\epsilon > 0$, existe $\delta > 0$ tal que $|f(\theta - t) - f(\theta)| < \frac{\epsilon}{3\|K\|_1}$ si $\|t\| < \delta$. En primer lugar veamos que $\int_{R^d} K_h(t) dt = 1$ ya que

$$\int_{R^d} K_h(t) dt = \int_{R^d} \frac{1}{h^d} K\left(\frac{t}{h}\right) dt = \int_{R^d} K(u) du = 1$$

Tomando la diferencia $|f_h(\theta) - f(\theta)| = \left| \int_{R^d} f(\theta - t) K_h(t) dt - f(\theta) \int_{R^d} K_h(t) dt \right| =$

$$\left| \int_{R^d} (f(\theta - t) - f(\theta)) K_h(t) dt \right| \leq \frac{\epsilon}{3\|K\|_1} \int_{\|t\| < \delta} |K_h(t)| dt + \int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt + \int_{\|t\| \geq \delta} |f(\theta)| |K_h(t)| dt$$

Además $\int_{\|t\| \geq \delta} |K_h(t)| dt < \frac{\epsilon}{3|f(\theta)|}$ si h es suficientemente chico por ser la cola de una integral convergente. Definamos la función $\ell : R^d \rightarrow R$ mediante $\ell(\theta) = |K(\theta)| \|\theta\|^d$. Por ii) vemos que $\ell(\theta) \rightarrow 0$ si $\|\theta\| \rightarrow +\infty$ luego

$$\int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt = \int_{\|t\| \geq \delta} |f(\theta - t)| \frac{\ell(t/h)}{\|t\|^d} dt \leq \frac{1}{\delta^d} \sup_{\|t\| \geq \delta} \ell(t/h) \int_{\|t\| \geq \delta} |f(\theta - t)| dt$$

Como $\sup_{\|t\| \geq \delta} \ell(t/h) \rightarrow 0$ si $h \rightarrow 0$ pues $\|t/h\| \rightarrow +\infty$ y $\int_{\|t\| \geq \delta} |f(\theta - t)| dt \leq \|f\|_1$ entonces $\int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt < \frac{\epsilon}{3}$ luego obtenemos que $|f_h(\theta) - f(\theta)| < \epsilon$.

Usando el teorema anterior obtenemos resultados similares a los del caso unidimensional. Debe hacerse una serie de elecciones, en principio el núcleo d dimensional y luego la matriz de anchos de banda, H lo que implica dos elecciones la parametrización y la elección de los h .

Planteemos las siguientes hipótesis:

H1) Todas las componentes de la matriz Hessiana de f (\mathfrak{H}_f) son continuas a trozos y de cuadrado integrable.

H2) $H = H_N$ es una sucesión de matrices tales que todos los coeficientes de H tienden a 0 cuando $N \rightarrow \infty$ y $\lim_{N \rightarrow \infty} N |H| = \infty$. La razón entre el mayor valor propio y el menor valor propio de H permanece acotado para todo n .

H3) K es un núcleo d dimensional acotado de soporte compacto y cumple $\int z K(z) dz = 0$, y $\int z K(z) dz = 0$, $\int z z^t K(z) dz = \mu_2(K) I_d$ con $\mu_2(K) = \int z_i^2 K(z) dz$ para todo i .

Bajo estas condiciones se obtiene aplicando el desarrollo de Taylor multidimensional al igual que en el caso unidimensional (al que vimos en detalle) que:

$$E(f_{N,H}(\theta)) = \int K_H(z - \theta) f(z) dz = \int K_H(s) f(\theta + Hs) ds \approx \int K_H(s) \{f(\theta) + s^t H^t \nabla f(\theta) + \frac{1}{2} s^t H^t \mathfrak{H}_f \nabla H s\} ds$$

$$E(f_{N,H}(\theta)) \approx f(\theta) + \frac{1}{2} \mu_2(K) \text{tr}(H^t \mathfrak{H}_f(\theta) H)$$

$$V(f_{N,H}) = \frac{1}{N} \int (K_H(z - \theta))^2 dz - \frac{1}{N} (E(f_{N,H}(\theta)))^2 \approx \frac{1}{N|H|} \int (K_H(s))^2 f(\theta + Hs) ds$$

$$V(f_{N,H}) \approx \frac{1}{N|H|} \int (K_H(s))^2 \{f(\theta) + s^t H^t \nabla f(\theta)\} ds \approx \frac{1}{N|H|} \|K\|_2^2 f(\theta)$$

El sesgo integrado se define como $SI(H) = \int (E(f_{N,H}(\theta)) - f(\theta))^2 d\theta$ y definimos el sesgo integrado asintótico $SIA(H)$ como el término de primer orden de $SI(H)$ o sea $\frac{SI(H) - SIA(H)}{SIA(H)} = o(1)$ cuando $N \rightarrow \infty$, $H \rightarrow 0$ y $NH \rightarrow \infty$.

La varianza integrada se define como $VI(H) = \int (E(f_{N,H}(\theta)) - E(f_{N,H}(\theta)))^2 d\theta$ e igualmente definimos la varianza integrada asintótica. El error cuadrático integrado medio asintótico (ECIMA) se define $ECIMA(H) = SIA(H) + VIA(H)$ Tenemos a partir de lo anterior que:

$$SIA(H) = \frac{1}{4} \mu_2(K)^2 \int \text{tr}^2(H^t \mathfrak{H}_f(\theta) H) d\theta$$

$$VIA(H) = \frac{1}{N|H|} \|K\|_2^2$$

$$ECIMA(f_{N,H}) = \frac{1}{4} \mu_2(K)^2 \int \text{tr}^2(H^t \mathfrak{H}_f(\theta) H) d\theta + \frac{1}{N|H|} \|K\|_2^2$$

En el caso particular $H = hI$ se obtiene

$$ECIMA(f_{N,H}) = \frac{\int K(z)^2 dz}{N h^d} + \frac{1}{4} h^4 \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta$$

En este caso se puede obtener una fórmula explícita para la ventana óptima que da

$$(3.15) \quad h = \left(\frac{d \int K(z)^2 dz}{N \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta} \right)^{1/(d+4)}$$

Si calculamos

$$\inf_{h>0} ECIMA(f_{N,H}) = \frac{d+4}{4d} (\mu_2(K)^{2d} d^4 (\int K(z)^2 dz)^4 \int (\nabla^2 f(\theta))^2 d\theta)^d N^{-4})^{1/(d+4)}$$

o sea, el orden de convergencia es $N^{-4/(d+4)}$ el cual disminuye a medida que crece la dimensión, justamente este orden manifiesta claramente la llamada "maldición de la dimensionalidad".

Distribución asintótica del estimador

Estos estimadores adecuadamente normalizados convergen a la distribución normal. Comencemos dando sin demostración la conocida desigualdad de Jensen.

Teorema 3.3.6. Desigualdad de Jensen

Sea μ una medida positiva sobre una σ -álgebra M en un conjunto Ω tal que $\mu(\Omega) = 1$. Sea f una función real en $L^1(\mu)$ con $a < f(x) < b$ para todo $x \in \Omega$. Si φ es convexa en (a,b) entonces

$$\varphi\left(\int_{\Omega} f d\mu\right) \leq \int_{\Omega} \varphi(f) d\mu$$

Lema 3.3.7. Suponiendo que $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$ y bajo las condiciones del lema 3.3.1.

$$\frac{f_N(\theta) - E f_N(\theta)}{\sigma(f_N(\theta))} \xrightarrow{w} N(0, 1)$$

Demostración: El estimador de la densidad de la habilidad lo podemos escribir como

$$f_N(\theta) = \frac{1}{Nh} \sum_{k=1}^N (K \frac{X - \Theta_k}{h}) = \frac{1}{N} \sum_{k=1}^N W_{N_k} \text{ con } W_{N_k} = \frac{1}{h} K\left(\frac{X - \Theta_k}{h}\right)$$

O sea, se tiene un sistema triangular de variables idénticamente distribuidas por filas. Utilizaremos la condición de Liapunov para la convergencia a la Normal, de más está decir que esta condición implica las de Lindeberg. Una condición suficiente para que:

$$\frac{\hat{f}_N(\theta) - E \hat{f}_N(\theta)}{\sigma(\hat{f}_N(\theta))} \xrightarrow{w} Z \sim N(0, 1) \text{ es que } \frac{E(|W_{N_1} - E(W_{N_1})|^{2+\delta})}{N^{\delta/2} \sigma(W_{N_1})^{2+\delta}} \longrightarrow 0 \text{ con } N_1 \rightarrow \infty \text{ para algún } \delta > 0$$

$$E(|W_{N_1}|^{2+\delta}) = \frac{1}{h^{2+\delta}} E(K^{2+\delta} \left(\frac{\theta - \Theta_1}{h}\right)) = \frac{1}{h^{1+\delta}} \frac{1}{h} \int K^{2+\delta} \left(\frac{\theta - t}{h}\right) f(t) dt$$

Si se cumplen las hipótesis del lema 3.3.1 $\frac{1}{h} \int K^{2+\delta} \left(\frac{\theta - t}{h}\right) f(t) dt \rightarrow f(\theta) \int K^{2+\delta}(t) dt$ así pues,

$$h^{1+\delta} E(|W_{N_1}|^{2+\delta}) \rightarrow f(\theta) \int K^{2+\delta}(t) dt$$

De igual modo: $\sigma^2(W_{N_1}) = \frac{1}{h^2} Var \left(K \left(\frac{\theta - \Theta_1}{h}\right)\right) = \frac{1}{h^2} \int K^2 \left(\frac{\theta - t}{h}\right) f(t) dt$.

Como $\frac{1}{h} \int K^2 \left(\frac{\theta - t}{h}\right) f(t) dt$ se tiene que: $h \sigma^2(W_{N_1}) \rightarrow f(\theta) \int K^2(t) dt$

También $E(W_{N_1}) \rightarrow f(\theta)$ si $\int K(u) du = 1$

$$E(|W_{N_1} - E(W_{N_1})|^{2+\delta}) \leq E(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta}$$

usando la desigualdad: $\left(\frac{a+b}{2}\right)^{2+\delta} \leq \frac{a^{2+\delta} + b^{2+\delta}}{2}$ obtenemos:

$$(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta} \leq \frac{2^{2+\delta}}{2} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta}) = 2^{1+\delta} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta})$$

Aplicando la desigualdad de Jensen a la función $\varphi = \theta^{2+\delta}$ tenemos que:

$$(E|W_{N_1}|)^{2+\delta} \leq E(|W_{N_1}|^{2+\delta}) \text{ así pues}$$

$$E(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta} \leq E(2^{1+\delta} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta})) \leq$$

$$\leq E(2^{1+\delta}(|W_{N1}|^{2+\delta} + E(|W_{N1}|^{2+\delta}))) = 2^{1+\delta}(E|W_{N1}|^{2+\delta} + E|W_{N1}|^{2+\delta}) = 2^{2+\delta}E(|W_{N1}|^{2+\delta})$$

luego

$$\frac{E(|W_{N1} - E(W_{N1})|^{2+\delta})}{N^{\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}} \leq \frac{2^{2+\delta}E(|W_{N1}|^{2+\delta})}{N^{\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}} = \frac{2^{2+\delta}}{h^{1+\delta}}h^{1+\delta}E(|W_{N1}|^{2+\delta})/N^{\frac{\delta}{2}}\frac{h^{1+\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}}{h^{1+\frac{\delta}{2}}}$$

$$\leq \frac{C}{h^{1+\delta}}/\frac{N^{\delta/2}}{h^{1+\delta/2}} = \frac{C}{(Nh)^{\delta/2}} \rightarrow 0 \text{ si } Nh \rightarrow \infty \text{ pues}$$

$$h^{1+\delta}E(|W_{N1}|^{2+\delta}) \rightarrow f(\theta) \int K^2(t)dt \text{ y } h^{1+d/2}\sigma(w_{N1})^{2+\delta} \rightarrow cte$$

así pues se cumple la condición de Liapunov y hemos probado que:

$$\frac{f_N(\theta) - Ef_N(\theta)}{\sigma(f_N(\theta))} \xrightarrow{w} N(0, 1)$$

Teorema 3.3.8. Supongamos que $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$; que K es acotada y que $\int K(u)du = 1$, además se cumple que $\mu^2 K(u)$ es acotada, $Nh \rightarrow \infty$ y $Nh^3 \rightarrow 0$. También la densidad f es Lipchitz en un entorno de θ , entonces:

- a) $Var(\sqrt{Nh}\widehat{f_N}(\theta)) \rightarrow f(\theta) \int K^2(u)du$
- b) $\sqrt{Nh} \left(\frac{1}{Nh} \sum_{i=1}^N E(K(\frac{\theta - \Theta_i}{h})) - f(\theta) \right) \rightarrow 0$
- c) $\sqrt{Nh} (\widehat{f_N}(\theta) - f(\theta)) \rightarrow N(0, f(\theta) \int K^2(u)du)$

Demostración: a) Como K es acotado, $\int K(u)du = 1$ entonces $\int K^2(u)du < \infty$.

Sabemos que $|uK(u)| \rightarrow 0$ si $|u| \rightarrow \infty$ pues $u^2 K(u)$ acotado luego se cumplen las hipótesis del lema 3.3.1, así que a) se cumple,

- b) existe $\delta > 0$ tal que $|f(\theta + s) - f(\theta)| \leq c|s|$; $|s| < \delta$

$$|\sqrt{Nh} \left(\frac{1}{Nh} \sum_{i=1}^N E(K(\frac{\theta - \Theta_i}{h})) - f(\theta) \right)| = |\sqrt{Nh} \left(\int \frac{1}{h} K(\frac{\theta - t}{h}) f(t)dt - f(\theta) \right)| =$$

$$|\sqrt{Nh} \left(\int K(u)f(\theta + uh)du - f(\theta) \int K(u)du \right)| = |\sqrt{Nh} \int K(u)|f(\theta + uh) - f(\theta)|du| \leq$$

$$\leq \sqrt{Nh} \left[hC \int_{|hu| < \delta} |uK(u)|du + \int_{|u| > \delta/h} |K(u)||f(\theta + uh)|du + |f(\theta)| \int_{|u| > \delta/h} |K(u)|du \right] =$$

Sea $C_1 = \|u^2 K(u)\|_\infty$

$$= \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + \int_{|u| > \delta/h} \frac{|u^2 K(u)|}{|u^2|} f(\theta + uh)du + |f(\theta)| \int_{|u| > \delta/h} \frac{|u^2 K(u)|}{u^2} \right] \leq$$

$$\leq \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + \frac{h^2}{\delta^2} C_1 \int_{|u| > \delta/h} f(\theta + uh)du + C_1 f(\theta) \int_{|u| > \delta/h} \frac{du}{u^2} \right] \leq$$

$$\leq \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + h \frac{C_1}{\delta^2} \int_R f(t) + C_1 f(\theta) h \int_{|t| > \delta} \frac{dt}{t^2} \right] = \sqrt{Nh^3} [C_2 + C_3 + C_4] \rightarrow 0$$

$$\text{si } Nh^3 \rightarrow 0 \text{ pues } \int_{|u| > \delta/h} \frac{du}{u^2} = h \int_{|t| > \delta} \frac{dt}{t^2} \int_{|u| > \delta/h} f(\theta + uh)du = \int_R f(t) \frac{dt}{h} \text{ y } \int_{|hu| < \delta} |uK(u)|du$$

está acotado al igual que: $f(\theta) \int_{|t| > \delta} \frac{dt}{t^2}$

c) $\sqrt{Nh}(f_N(\theta) - f(\theta)) = \sqrt{Nh}(f_N(\theta) - Ef_N(\theta)) + \sqrt{Nh}[Ef_N(\theta) - f(\theta)]$
sabemos que: $\sqrt{Nh}[Ef_N(\theta) - f(\theta)] \rightarrow 0$ consideremos, $\sigma(f_N(\theta)) \frac{(f_N(\theta) - Ef_N(\theta))}{\sigma(f_N(\theta))}$
se tiene que: $\frac{(f_N(\theta) - Ef_N(\theta))}{\sigma(f_N(\theta))} \sim N(0, 1)$ además, $Nh\sigma^2(f_N(\theta)) \rightarrow f(\theta) \int K^2(u)du$
entonces: $\sqrt{Nh}\sigma(f_N(\theta)) \rightarrow (f(\theta) \int K^2(u)du)^{1/2}$ así pues,
 $\sqrt{Nh}(f_N(\theta) - f(\theta)) \rightarrow N(0, f(\theta) \int K^2(u)du)$

Convergencia completa del estimador

El estimador no solo converge en probabilidad o sea es debilmente consistente sino que también o hace completamente lo cual implica convergencia casi segura o sea la consistencia fuerte. Recordemos que una sucesión de variables aleatorias X_n converge completamente a otra variable aleatoria X si para todo $\varepsilon > 0$

$$\sum_{i=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$$

El teorema que prueba la convergencia completa utiliza una desigualdad exponencial (desigualdad de Bernstein) para acotar superiormente la serie de probabilidades mediante una serie convergente.

Teorema 3.3.9. Sea $f_N(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\Theta}_i}{h}\right)$, con $\hat{\Theta}_1, \dots, \hat{\Theta}_N$ estimadores de $\Theta_1, \dots, \Theta_N$ independientes y que tienen densidad f . Supongamos que el rasgo se mide sin error es decir $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$. El núcleo K cumple las hipótesis del lema 3.3.1, e $\int K = 1$

En estas condiciones el estimador basado en núcleos converge completamente a la densidad f en cada punto de continuidad de f .

Demostración: Queremos probar que $\sum_{i=1}^{\infty} P(|f_N(\theta) - f(\theta)| > \epsilon) < \infty$.

$$\{|f_N(\theta) - f(\theta)| > \epsilon\} = \left\{ \left| \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| > \epsilon \right\} = \left\{ \left| \frac{1}{N} \left[\sum_{i=1}^N \left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right) \right] \right| > \epsilon \right\}$$

$$\begin{aligned} \text{entonces } P(|f_N(\theta) - f(\theta)| > \epsilon) &= P\left(\left| \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right) \right| > \epsilon \right) \leq \\ &\leq P\left(\frac{1}{N} \sum_{i=1}^N \left| \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| > \epsilon \right) \end{aligned}$$

Consideremos la variable $Y_i = \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta)$. Claramente las Y_i son independientes si las Θ_i lo son:

$$E(Y_i) = E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right)\right) - f(\theta) = \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) f(t) dt - f(\theta) \rightarrow 0 \text{ pues } \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) f(t) dt \rightarrow f(\theta)$$

$$\text{Además, } |Y_i| = \left| \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| < C \text{ pues } K \text{ es acotada.}$$

$$E(Y_i^2) = E\left(\frac{1}{h^2} K^2\left(\frac{\theta - \Theta_i}{h}\right)\right) - 2f(\theta)E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right)\right) + f^2(\theta) =$$

$$\frac{1}{h} \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt - 2f(\theta) \frac{1}{h} \int K\left(\frac{x-t}{h}\right) f(t) dt + f^2(\theta) \rightarrow \frac{1}{h} f(\theta) \int K^2(t) dt - f^2(\theta)$$

$$\text{ya que } \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow f(\theta) \int K^2(t) dt \text{ y } (1/h) \int K\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow f(\theta)$$

Utilizando la desigualdad de Bernstein tenemos:

$$P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq 2 \exp\left(\frac{-N^2 \epsilon^2}{2\left(N\left(\frac{1}{h} f(\theta) \int K^2(t) dt - f^2(\theta)\right) + cN\epsilon\right)}\right)$$

Operando en el exponente

$$\frac{-N^2 \epsilon^2}{2\left(N\left(\frac{f(\theta)}{h} \int K^2(t) dt - f^2(\theta) + c\epsilon\right)\right)} = \frac{-N\epsilon^2}{2\left(\frac{C_2}{h} - C_1 + c\epsilon\right)} = \frac{-N\epsilon^2}{2\left(\frac{C_2}{h} + C_3\right)} \leq \frac{-N\epsilon^2}{2\frac{C_2}{h}} = -NhC'$$

$$\text{Así pues } P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq 2 \exp(-NhC') \text{ entonces } \sum P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq \sum 2e^{-NhC'}$$

$$\text{Si } Q(N) = \frac{Nh}{\log N} \text{ tenemos } Nh = Q(N) \log N \text{ entonces } 2e^{-C' \log N Q(N)} = 2e^{\log N^{-C' Q(N)}} = 2N^{-C' Q(N)}$$

Si $Q(N) \rightarrow \infty$ resulta que $\sum 2N^{-C' Q(N)}$ converge, así pues hay convergencia completa.

Estimación de la cópula

Una forma sencilla de estimar la cópula es considerar la cópula empírica. Supongamos que tenemos los estimados $(\widehat{\theta}_{1i}, \dots, \widehat{\theta}_{di})$, $i = 1, \dots, N$ de $(\theta_{1i}, \dots, \theta_{di})$, $i = 1, \dots, N$

Sabemos que si $(U_{1i}, \dots, U_{di}) = (F_1(\theta_{1i}), \dots, F_d(\theta_{di}))$, $i = 1, \dots, n$. entonces

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d]$$

Consideremos $(\widehat{U}_{1i}, \dots, \widehat{U}_{di}) = (F_1(\widehat{\theta}_{1i}), \dots, F_d(\widehat{\theta}_{di}))$, $i = 1, \dots, n$. Observemos que este vector coincide con los estimados originales de los rasgos en cubo $[0, 1]^d$. Luego el candidato a estimador de la cópula es

$$C(u_1, \widehat{u_2}, \dots, u_d) = \mathbb{P}[\widehat{U}_1 \leq u_1, \dots, \widehat{U}_d \leq u_d]$$

Este lo podemos calcular empíricamente como

$$(3.16) \quad C(u_1, \widehat{u_2}, \dots, u_d) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\widehat{U}_{1i} \leq u_1, \dots, \widehat{U}_{di} \leq u_d)$$

Es claro que falta estudiar el comportamiento estadístico de este estimador es decir si es consistente.

3.4. Núcleos

En esta sección queremos presentar los núcleos que pueden usarse para estimar los modelos. Primeramente veamos núcleos univariados comúnmente usados, pues a partir de estos suelen contruirse los multivariados.

Los núcleos más comunes son:

Uniforme	$K(u) = \frac{1}{2} 1_{(u \leq 1)}$
Triangular	$K(u) = (1 - u) 1_{(u \leq 1)}$
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2) 1_{(u \leq 1)}$
Quartic o Biweight	$K(u) = \frac{15}{16}(1 - u^2)^2 1_{(u \leq 1)}$
Triweight	$K(u) = \frac{35}{32}(1 - u^2)^3 1_{(u \leq 1)}$
Tricubo	$K(u) = \frac{70}{81}(1 - u ^3)^3 1_{\{ u \leq 1\}}$
Gaussiano	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Coseno	$K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) 1_{(u \leq 1)}$
Logístico	$K(u) = \frac{1}{e^u + 2 + e^{-u}}$
Silverman	$K(u) = \frac{1}{2} e^{-\frac{ u }{\sqrt{2}}} \cdot \sin\left(\frac{ u }{\sqrt{2}} + \frac{\pi}{4}\right)$
Sigmoideo	$K(u) = \frac{2}{\pi} \frac{1}{e^u + e^{-u}}$

Cuadro 3.1: Tipos de núcleo

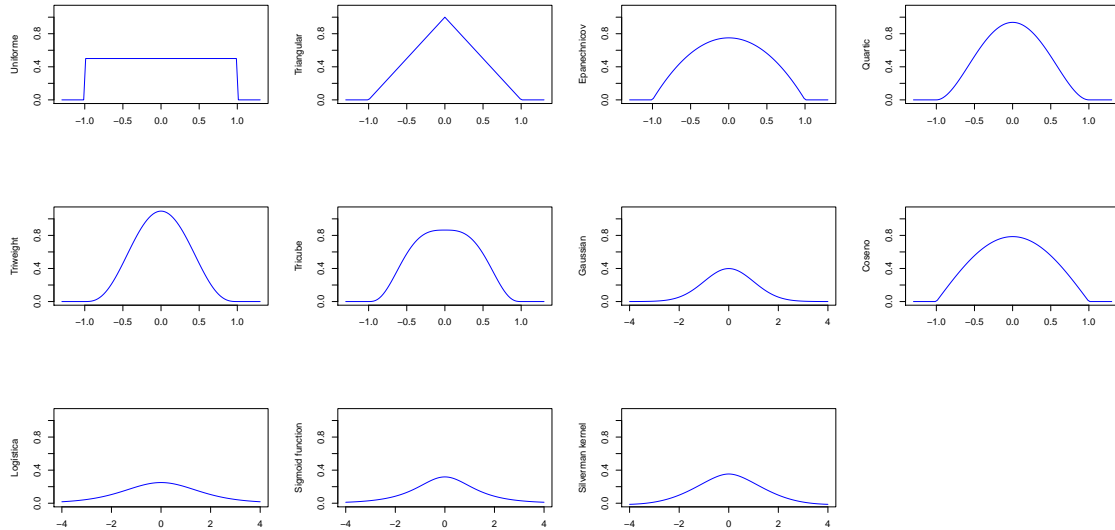


Figura 3.2: Principales núcleos unidimensionales

Hay dos formas clásicas de construir núcleos multivariados, la primera, es construyendo el llamado núcleo multiplicativo, y la segunda, es construir núcleos esféricos o simétricos radialmente.

El núcleo producto se crea mediante $K(u) = K_1(u_1)K_2(u_2)\dots K_d(u_d)$ donde cada K_i es un núcleo unidimensional. Como ejemplo de este tipo de núcleo tenemos el llamado núcleo multiplicativo de Epanechnikov $(\frac{3}{4})^d(1 - u_1^2)(1 - u_2^2)\dots(1 - u_d^2)1_{(|u_1| \leq 1)}\dots 1_{(|u_d| \leq 1)}$

Los núcleos radiales simétricos son de la forma $K^*(u) = \frac{K(\|u\|)}{\int_R K(t)}$ con $\|u\| = u^T u$ y K un núcleo unidimensional.

Por ejemplo el núcleo de Epanechnikov esférico $K(u) = \propto (1 - u^T u)1_{(u^T u \leq 1)}$ y el núcleo normal multivariado es: $K_{\mathbf{H}}(\mathbf{u}) = (2\pi)^{-d/2} |\mathbf{H}|^{-1/2} e^{-\frac{1}{2} \mathbf{u}^T \mathbf{H}^{-1} \mathbf{u}}$ donde \mathbf{H} es la matriz de varianzas covarianzas.

También está el núcleo de Bartlett-Epanechnikov donde si $c_d = \frac{\pi^{d/2}}{\Gamma((d/2)+1)}$ es el volumen de la esfera unitaria en el espacio R^d viene dado por $K(x) = \frac{(1-x^t x)(d+2)}{2c_d} 1_{(x^t x \leq 1)}$

Otros núcleos a ser usados son $K_2(x) = 3\pi^{-1}(1-x^t x)^2 1_{(x^t x \leq 1)}$ y $K_3(x) = 4\pi^{-1}(1-x^t x)^3 1_{(x^t x \leq 1)}$

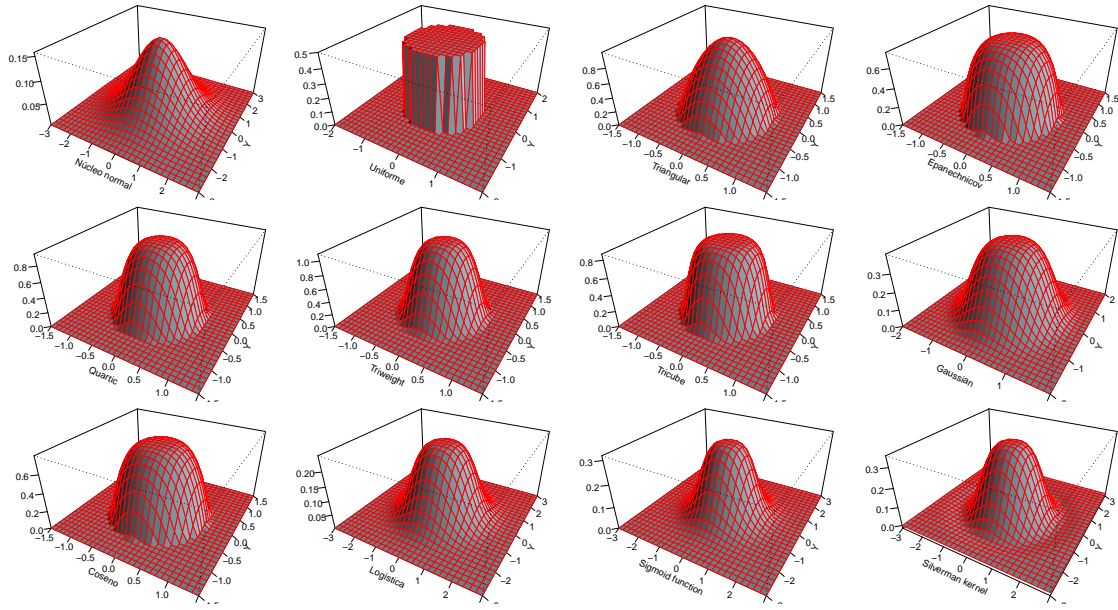


Figura 3.3: Núcleos multiplicativos de dimensión 2

Cabe preguntarse que tipo de núcleo elegir (multiplicativo o esférico) si queremos estimar la densidad del rasgo. Se puede probar que para comparar dos núcleos multivariados las magnitudes a comparar son $(\mu_2(K)^{2d}(\int K(z)^2 dz)^4)^{1/(d+4)}$. Entonces tomando el cociente de estas cantidades para el núcleo esférico sobre el núcleo multiplicativo podemos medir sus eficiencias relativas. Wand y Jones (1995) realizan estas comparación para estos núcleos basados en el núcleo unidimensional beta o sea $K(x) \propto (1-x^2)^p 1_{|x|<1}$.

p	d=2	d=3	d=4
0	0.955	0.888	0.811
1	0.982	0.953	0.916
2	0.983	0.953	0.915
3	0.984	0.956	0.919

Cuadro 3.2: Núcleo beta

Se observa que el núcleo esférico tiene un comportamiento asintótico mejor que el multiplicativo, siendo en particular muy importante para el núcleo uniforme ($p=0$) y que esta diferencia crece con la dimensión. Si queremos encontrar el núcleo óptimo en este sentido se debe encontrar el núcleo que perteneciendo a una familia minimice la cantidad $(\mu_2(K)^{2d}(\int K(z)^2 dz)^4)^{1/(d+4)}$. El núcleo multiplicativo que minimiza esta expresión es $K(x) = (\frac{3}{4})^d \prod_{i=1}^d (1-x_i^2) 1_{|x|<1}$ en el caso esférico es $K(x) = \frac{d\Gamma(d/2)}{4\pi^{d/2}}(d+2)(1-x^t x) 1_{x^t x \leq 1}$.

3.5. Elección del ancho de banda para densidad

Elegir el ancho de banda H a utilizar es un tema importante para la estimación, tanto de la densidad como de la CCI, debido a su efecto compensatorio entre el sesgo y la varianza. En una o dos dimensiones se puede elegir el ancho de banda apropiado observando sucesivas estimaciones para diferentes anchos de banda. Cuando queremos utilizar un método para elegir el ancho de banda suelen usarse dos enfoques: el método plug-in, en particular la llamada “regla del pulgar” y métodos de remuestreo, como la validación cruzada y el bootstrap.

Härdle y Müller (1997) introducen generalizaciones a la “regla del pulgar” de Silverman y a la validación cruzada por mínimos cuadrados.

Método plug-in de Härdle y Müller

Supongamos que el núcleo K es normal típico con dimensión d , luego se tiene $\|K\|_2^2 = 2^{-d}\pi^{-d/2}$; $\mu_2(K) = 1$ y

$$\int [tr \{H^T \mathfrak{H}_f(\theta) H\}]^2 dt = \frac{1}{2^{d+2}\pi^{d/2} |\Sigma|^{1/2}} \left[2tr(H^T \Sigma^{-1} H)^2 + \{tr(H^T \Sigma^{-1} H)\}^2 \right]$$

Se puede derivar fácilmente la fórmula de la “regla del pulgar” suponiendo diferentes condiciones sobre H y Σ .

En 3.15 se mostro que

$$(3.17) \quad h = \left(\frac{d \int K(z)^2 dz}{N \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta} \right)^{1/(d+4)}$$

Si los datos provienen de una distribución normal multivariante se obtiene que

$$h = Cn^{-1/(d+4)}$$

donde la constante C depende del núcleo.

Si el núcleo es gaussiano de dimensión d $C = \frac{4}{d+2}^{1/(d+4)}$, si el núcleo es K_2 , $C = 2.78$ y K_3 , $C = 3.12$, en tanto si se usa el núcleo de Epanechnikov de dimensión d , $C = (8c_d^{-1}(d+2)2^d\pi^{d/2})^{1/(d+4)}$

Si H y Σ son matrices diagonales $H = diag(h_1, \dots, h_d)$ y $\Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)$, entonces siguiendo el desarrollo de Silverman (1986) se obtiene $h_j = \left(\frac{4}{d+2} \right)^{1/(d+4)} N^{-1/(d+4)} \sigma_j$

Si se reemplaza σ_j por su estimación y el primer factor por 1 se tiene la regla de Scott (1992), $h_j = \left(\frac{4}{d+2} \right)^{1/(d+4)} N^{-1/(d+4)} \hat{\sigma}_j$.

Scott (1992) también halla para el caso del núcleo producto de dos normales y los datos normales bivariadas $h_i = \sigma_i(1 - \rho^2)^{5/12}(1 + \rho^2/2)^{-1/6}n^{-1/6}$ con $i = 1, 2$

Es difícil derivar la regla para el caso general de H y Σ , pero tomando la matriz de ventanas proporcional a la raíz de la matriz de covarianzas se generaliza la fórmula de Scott obteniendo $\hat{H} = N^{-1/(d+4)} \hat{\Sigma}^{1/2}$. Si se desea usar un núcleo no Gaussiano, consideramos un núcleo K y las funciones equivalentes $K_\delta = \delta^{-1}K(\bullet/\delta)$ con $\delta \geq 0$. Tenemos que $\|K_\delta\|_2^2 = \delta^{-d}\|K\|_2^2$ y $\mu_2(K_\delta) = \delta^2\mu_2(K)$. Se elige δ de tal manera que la relación entre la varianza y el sesgo en el $ECIMA(H, K_\delta)$ sea independiente de K_δ . Esto se cumple si

$$\mu_2^2(K_{\delta_0}) = \|K_\delta\|_2^2 \iff \delta_0 \left\{ \frac{\|K_\delta\|_2^2}{\mu_2^2(K)} \right\}^{1/(d+4)}$$

δ_0 se llama ancho de banda canónico del núcleo K . Si queremos calcular el ancho de banda para diferentes núcleos Härdle y Müller (1997) deducen la fórmula $H_B = \frac{\delta_0^B}{\delta_0^A} H_A$ donde δ_0^A y δ_0^B son los anchos de banda canónicos para los núcleos K^A y K^B y siendo H_A y H_B las matrices usadas con K^A y K^B respectivamente. Con esta elección se obtiene que

$$ECIMA(H_A, K^A) \approx ECIMA(H_B, K^B)$$

En Wand y Jones (1994) se encuentran detallados otros métodos plug-in.

Método de validación cruzada

El método de validación cruzada por mínimos cuadrados tiene por objeto encontrar el ancho de banda óptimo para el ECI. Tenemos que

$$ECI(H) = \int \{\hat{f}_H(t) - f(t)\}^2 dt = \int \hat{f}_H^2(t) dt - 2 \int \{\hat{f}_H(t) f(t)\} dt + \int f^2(t) dt$$

Ignoramos el último término de la ecuación, el primer término puede calcularse a partir de los datos. Para el segundo término puede ser estimado observemos que $\int \hat{f}_H(t) f(t) dt = E\hat{f}_H(T)$, donde ahora T es d -dimensional y puede estimarse mediante

$$E\hat{f}_H(T) = \frac{1}{n} \sum_{i=1}^n \hat{f}_{H,-1}(T_i) \text{ donde } \hat{f}_{H,-i}(t) = \frac{1}{n-1} \sum_{i \neq j=1}^n K_H(T_j - t)$$

Entonces tenemos el criterio de validación cruzada multivariada como:

$$CV(H) = \frac{1}{n^2 \det(H)} \sum_{i=j}^n \sum_{i=j}^n K \star K \{H^{-1}(T_j - t_i)\} - \frac{2}{n(n-1)} \sum_{i=j}^n \sum_{i=jj \neq i}^n K_H(T_j - T_i)$$

La dificultad viene porque el ancho de banda es ahora una matriz H $d \times d$ lo que implica tenemos que minimizar en $d(d+1)/2$ parámetros. Si se asume H como una matriz diagonal, esto se convierte en un problema de optimización d -dimensional.

3.6. Consistencia conjunta del modelo multivariado

Esta sección está dedicado a encontrar las condiciones que deben cumplir los estadísticos para obtener estimaciones consistentes del rasgo y la relación entre el tamaño muestral N y la cantidad de ítems n para obtener estimaciones consistentes de las CCI. A nuestro juicio esta sección es de capital importancia para justificar el uso del estimador.

Como antecedente tenemos a Douglas (1997) que demuestra la consistencia conjunta para el caso unidimensional. Douglas asume ciertas hipótesis sobre los ítems, las CCI, los tamaños muestrales N_n , los núcleos y los anchos de banda h_n . Considera 10 hipótesis que enumeramos a continuación:

1. Independencia local y unidimensionalidad.
2. Θ tiene distribución uniforme $(0,1)$.
3. \overline{P}'_n existe en $(0,1)$ y existe un número m tal que $0 < m < \overline{P}'_n(\theta)$ para todo $\theta \in (0,1)$.
4. Existe $C > 0$ tal que $np_{nk} < C$ para todo k donde $p_{nk} = P(\overline{U}_n = k)$.
5. Existen constantes M_L y M_U tal que $M_L n^{3/2} < N_n < M_U n^r$.

6. $P'_{n,i}$ existe y es continua en $(0,1)$ para todo n,i .
7. Para cada intervalo compacto $[a,b] \subset (0,1)$ existe una constante $M_{a,b}$ tal que para todo $\theta \in [a,b]$ y todo n, i $|P'_{n,i}(\theta)| < M_{a,b}$.
8. K tiene soporte acotado, en particular existe una constante C_K , tal que $K(x) > 0$, cuando $|x| < C_K$ y $K(x) = 0$, cuando $|x| > C_K$.
9. K es Lipschitziana de constante L_K .
10. Existe un número $\alpha \in (0, 1/2)$ y números positivos L_α y U_α tal que para todo n ,
$$L_\alpha n^{-\alpha} < h_n < U_\alpha n^{-\alpha}$$

A partir de estas hipótesis se obtiene primeramente:

Teorema 3.6.1. Dadas las hipótesis A1 a A7 $\max_{j=1,\dots,J_n} |\hat{\Theta}_{n,j} - \Theta_{n,j}| \rightarrow 0$ con probabilidad 1.

También se obtiene un resultado sobre el error cuadrático.

Teorema 3.6.2. Dadas las hipótesis 1 a 10 para $\theta_0 \in (0,1)$ fijo

$$(3.18) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(n^{-2\alpha})$$

donde α es la constante de la hipótesis 10.

Definiendo $\delta_{n,a,b} = \sup_{\theta \in (a,b)} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|$ se obtiene:

Teorema 3.6.3. Dadas las hipótesis 1 a 10, $\delta_{n,a,b} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

Finalmente prueba la consistencia conjunta pues a partir de los teoremas 3.6.1 y 3.6.3, y notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

Teorema 3.6.4. $\max\{\Delta_n, \delta_{n,a,b}\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

3.6.1. Preliminares

La consistencia conjunta requiere que el tamaño muestral y la longitud del test crezcan juntos.

Esto puede ejemplificarse mediante arreglos triangulares. Consideremos, primero, un test de n ítems, administrado a N_n examinados, elegidos al azar. Se estiman las CCI de los n ítems y los N_n rasgos. Entonces ahora, se considera un nuevo test de $n+1$ ítems, que no necesariamente contengan los n anteriores, y se administra a N_{n+1} examinados.

Haremos la estimación para cada fila en el arreglo triangular así obtenido y estudiaremos la adecuación de las CCI y de los estimados del rasgo a medida que n tiende a infinito.

Tenemos entonces la sucesión de test:

$$P_{n,1}, P_{n,2}, \dots, P_{n,n}$$

$$P_{n+1,1}, P_{n+1,2}, \dots, P_{n+1,n}, P_{n+1,n+1}$$

$$P_{n+2,1}, P_{n+2,2}, \dots, P_{n+2,n+1}, P_{n+2,n+2}$$

.....

.....

y la sucesión de rasgos:

$$\theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,N_n}$$

$$\theta_{n+1,1}, \theta_{n+1,2}, \dots, \theta_{n+1,N_n}, \theta_{n+1,N_{n+1}}$$

$$\theta_{n+2,1}, \theta_{n+2,2}, \dots, \theta_{n+2,N_{n+1}}, \theta_{n+2,N_{n+2}}$$

.....

.....

Se quiere entonces encontrar las condiciones para que la sucesión de estimadores de las curvas converjan a las verdaderas CCI y la sucesión de rasgos converja a los rasgos verdaderos conjuntamente.

Se obtiene bajo condiciones generales un resultado para la consistencia del rasgo donde si se cumplen las hipótesis 1) a 7) (ver la sección de hipótesis) $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad.

Si, además, se cumple 8), entonces, $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente (en particular con probabilidad 1).

Se encuentra también una desigualdad para el error cuadrático medio Se prueba que cuando se cumplen las hipótesis 1) a 12) y para $\theta_0 \in (0, 1)^d$ fijo

$$(3.19) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

El tercer resultado importante es sobre la consistencia de la estimación de las CCI. Dadas las hipótesis 1) a 12): $\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$ si $n \rightarrow +\infty$ completamente con $(a,b)^d \subset (0, 1)^d$.

Finalmente, se obtiene la consistencia conjunta de los rasgos y las CCI. Bajo las hipótesis 1) a 12): $\max\{\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

3.6.2. Sucesión de funciones asintóticamente Lipschitz

Una de las condiciones encontradas sobre la sucesión de las funciones de distribución de los estimadores de los rasgos es lo que llamaremos que estas deben ser asintóticamente Lipschitz.

Definición 3.6.5. Diremos que una sucesión de funciones $f_n : V \rightarrow R^d$ con $V \subset R^k$ acotado es asintóticamente Lipschitz si existen sucesiones L_n y $\gamma_n > 0$ tales que: existen $\delta > 0$ y $\lambda > 0$ tal que $\delta < L_n < \lambda$ para todo n y $\gamma_n \rightarrow 0$ y $\|f_n(x) - f_n(y)\| \leq L_n \|x - y\| + \gamma_n \forall x, y \in V$.

Esta definición está motivada por la necesidad de aproximarse al concepto de funciones Lipschitz cuando se tienen sucesiones de funciones de distribución de variables discretas o distribuciones empíricas. Es claro que estas funciones al ser escalonadas no pueden ser Lipschitz pero pueden cumplir la desigualdad anterior que es muy útil. A partir de la definición si la sucesión de funciones tiene límite, entonces, a partir de la continuidad de la norma el límite es una función lipschitziana. Veamos algunos ejemplos para clarificar.

Ejemplo 3.6.6. Supongamos que tenemos n variables aleatorias Bernoulli X_i , y consideremos la media muestral $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$. Si para algún $K > 0$, $nP(\sum_{i=1}^n X_i = k) < K$, $k = 0, 1, \dots, n$, (esta

condición implica que el histograma es uniformemente acotado), entonces para $0 \leq x \leq x' \leq 1$, la sucesión de funciones de distribución de \bar{X} son asintóticamente Lipschitz.

$$\begin{aligned} F_n(x') - F_n(x) &= \sum_{k=[nx]+1}^{[nx']} P(\sum_{i=1}^n X_i = k) \leq \sum_{k=[nx]+1}^{[nx']} \frac{K}{n} = \\ &= \frac{K}{n}([nx'] - [nx]) \leq \frac{K}{n}(n(x' - x) + 1) = K[(x' - x) + \frac{1}{n}] = K(x' - x) + \frac{K}{n} \end{aligned}$$

Ejemplo 3.6.7. Consideremos ahora como estimador la media ponderada $\bar{X} = \sum_{i=1}^n \omega_i X_i$.

Representemos por $(X_1^j, \dots, X_i^j, \dots, X_n^j)$ $j = 1, \dots, 2^n$ las posibles enuplas de valores que pueden tomarse y sea $I_n = \min_{j \neq k} | \sum_{i,k} \omega_i (X_i^j - X_i^k) |$. Si $\frac{P(\sum \omega_i X_i = k)}{I_n} \leq K$ entonces,

$$F_n(x') - F_n(x) = \sum_{k \in (x, x']} P(\sum \omega_i X_i = k) \leq K I_n (\frac{x' - x}{I_n} + 1) = K(x' - x) + I_n.$$

Luego, si $I_n \rightarrow 0$ la sucesión de funciones de distribución son asintóticamente de Lipschitz.

Ejemplo 3.6.8. Observemos también que si X_n es una sucesión de variables discretas con función de distribución F_n y si existen sucesiones $f(n) > 0$, $g(n) > 0$ y $h(n) > 0$ y constantes K_1 y K_2 tales que, $\sup_k \{P(X_n = k)\} \leq \frac{K_1}{f(n)}$ con $k \in \text{Rec}\{X_n\}$, para todo $0 \leq x \leq x' \leq 1$

$$\sum_{k \in \text{Rec}\{X_n\}} \chi_{(x, x']}(k) \leq K_2 g(n)(x' - x) + h(n)$$

con $\frac{h(n)}{f(n)} \rightarrow 0$ y para algún $\delta > 0$ y $\lambda > 0$; $\delta < \frac{g(n)}{f(n)} < \lambda$ para todo n entonces:

$$F_n(x') - F_n(x) \leq \sup_k \{P(X_n = k)\} \sum_{k \in \text{Rec}\{X_n\}} \chi_{(x, x']}(k) \leq \frac{K_1 K_2 g(n)}{f(n)}(x' - x) + \frac{h(n)}{f(n)}$$

lo que implica que F_n son asintóticamente lipschitzianas.

Como un primer resultados se tiene el siguiente lema:

Lema 3.6.9. Consideremos una sucesión de estadísticos X_n con función de distribución F_n , que converge completamente a F , la cual es lipschitziana de constante L . Entonces F_n es asintóticamente Lipschitz.

Demostración: Consideremos $0 \leq x \leq x' \leq 1$.

$$|F_n(x') - F_n(x)| \leq |F_n(x') - F(x')| + |F(x') - F(x)| + |F(x) - F_n(x)| \leq L \|x - x'\| + 2 \sup_x |F_n(x) - F(x)|$$

Como $|F_n(x) - F(x)|$ está acotado para todo x , notemos por $\gamma_n = \sup_x |F_n(x) - F(x)|$, entonces, $|F_n(x') - F_n(x)| \leq L \|x - x'\| + 2\gamma_n$ donde $\gamma_n \rightarrow 0$ pues $F_n \rightarrow F$ completamente.

Corolario 3.6.10. Si tomamos como estimador la media ponderada $\bar{X} = \sum_{i=1}^n \omega_i X_i$ donde las X_i son variables aleatorias independientes idénticamente distribuidas (iid).

Sea $a_n = \sum_{i=1}^n \omega_i^2$. Utilizando el resultado de Fisher (1992) tenemos que si $\limsup_{t \rightarrow \infty} \frac{1}{t} \#\{n : \frac{a_n}{\omega_n^2} < t\} < \infty$, entonces $\frac{S_n}{\sqrt{a_n}}$ converge en distribución a la normal típica de donde las distribuciones de $\sum_{i=1}^n \frac{\omega_i}{\sqrt{a_n}} X_i$ son asintóticamente Lipschitz.

3.7. Hipótesis para la consistencia conjunta

Presentamos a continuación, a los efectos de simplificar la lectura, las hipótesis necesarias para obtener la consistencia conjunta de los estimadores de los rasgos y de las CCI.

1) Independencia Local.

2) El rasgo Θ tiene distribuciones marginales uniformes $U[0, 1]$.

3) Para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2 \text{ para } 1 \leq l \leq d$$

4) Las funciones $F_{n,l} = P(g_{n,l}(X) \leq x)$ son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$

$$5) \frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0.$$

$$6) \text{ Para } 0 < \alpha < 1/2, \frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$$

$$7) H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}} < N < Mn^r \text{ con } r > d+1$$

$$8) \lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$$

9) El núcleo K es simétrico, acotado y con soporte compacto.

10) Las CCI tienen derivadas parciales continuas en $(0,1)^d$ y además en cada compacto $[a,b]^d \subset (0,1)^d$ y todo par (n,i) están uniformemente acotadas.

$$11) \text{ Para } 0 < \alpha < 1/2, \max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$$

12) El núcleo K es Lipschitz de constante J.

3.8. Prueba de la consistencia conjunta

Antes de continuar, veamos algunos resultados sobre desigualdades de concentración, que daremos sin demostración y nos ayudarán en lo que sigue. Para una descripción detallada de estos resultados recomendamos leer el excelente trabajo de Boucheron, Lugosi y Massart (2003).

Consideremos un conjunto \mathcal{X} y $g : \mathcal{X}^n \rightarrow R$ una función medible y sea $Z = g(X_1, \dots, X_n)$ donde X_1, \dots, X_n son variables aleatorias independientes (no tienen por qué ser idénticamente distribuidas) que toman valores en \mathcal{X} . Consideremos $E(Z)$ y $E_i(Z) = E(Z | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$.

Sea X'_1, \dots, X'_n una copia independiente de X_1, \dots, X_n y $Z'_i = g(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$.

Se obtienen unos primeros resultados:

Teorema 3.8.1.

$$V(Z) \leq \sum_{i=1}^n E[(Z - E_i(Z))^2]$$

Teorema 3.8.2. Desigualdad de Efron-Stein.

$$V(Z) \leq \frac{1}{2} \sum_{i=1}^n E[(Z - Z'_i)^2]$$

Definición 3.8.3. Diremos que una función $g : \mathcal{X}^n \rightarrow R$ tiene la propiedad de diferencias acotadas si existen constantes no negativas c_1, \dots, c_n , tales que,

$$\sup_{x_1, \dots, x_n, x'_i \in \mathcal{X}} |g(x_1, \dots, x_i, \dots, x_n) - g(x_1, \dots, x'_i, \dots, x_n)| \leq c_i \text{ para } 1 \leq c_i \leq n.$$

Para funciones de diferencias acotadas se tiene el siguiente corolario:

Corolario 3.8.4. Si g tiene la propiedad de diferencias acotadas con constantes c_1, \dots, c_n

$$V(Z) \leq \frac{1}{2} \sum_{i=1}^n c_i^2$$

A partir de la desigualdad logarítmica simétrica de Sobolev, debida a Massart (2000), se prueba que:

Teorema 3.8.5. Si existe una constante C positiva, tal que, casi seguramente

$$\sum_{i=1}^n (Z - Z'_i)^2 \leq C$$

entonces para todo $t > 0$, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4C}}$

Para las funciones de diferencias acotadas se tiene el corolario:

Corolario 3.8.6. Si g tiene la propiedad de diferencias acotadas con constantes c_1, \dots, c_n entonces, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^n c_i^2}}$.

Esta cota puede ser mejorada y McDiarmid (1989) prueba que, bajo las condiciones del corolario 3.8.6, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^n c_i^2}}$

En realidad, tampoco es necesario que g sea de diferencias acotadas, pues se puede aplicar el teorema y obtenemos una condición mucho más débil.

Corolario 3.8.7. Si para g existen constantes c_1, \dots, c_n tales que:

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g(x_1, \dots, x_i, \dots, x_n) - g(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_i^2$$

$$P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^n c_i^2}}$$

Consideremos de ahora en adelante en R^d la norma $\|x\| = \max_{i=1 \dots d} |x_i|$

Lema 3.8.8. Sea $G_{g_n}(x)$ definida por 3.1. Supongamos que para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$(3.20) \quad \sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2 \text{ para } 1 \leq l \leq d$$

Supongamos, además, que las $F_{n,l}$ correspondientes son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$. Entonces para $\epsilon > 0$:

$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$. En particular si $c_n = \max_{1 \leq l \leq d, 1 \leq i \leq n} c_{n,l,i}$

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq 2de^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 nc_n^2}}$$

Observación: En el caso unidimensional $g_n(x_1, \dots, x_n) = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$. Ahora $|g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)| = \frac{1}{n}$, entonces,

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n \frac{1}{n^2} = \frac{1}{n} \text{ luego,}$$

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) = P(\|G_{g_n}(\bar{X}) - G_{g_n}(E(\bar{X}))\| > \epsilon) \leq 2e^{-\frac{(\epsilon - \frac{K}{n})^2}{4K^2 n \frac{1}{n^2}}} = 2e^{-\frac{n(\epsilon - \frac{K}{n})^2}{4K^2}}$$

$$\begin{aligned} \text{Demostración: } P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) &\leq P(\|g_n(X) - E(g_n(X))\| > \frac{\epsilon - \gamma_n}{L_n}) = \\ &= P(\max_{1 \leq l \leq d} |g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) = P(\bigcup_{1 \leq l \leq d} |g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) \leq \\ &\leq \sum_{l=1}^d P(|g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) = \end{aligned}$$

$$= \sum_{l=1}^d \int_{[0,1]^d} P(|g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n} \mid \Theta = \theta) f_\theta(\theta) d\theta \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$$

Si $c_n = \max_{1 \leq l \leq d, 1 \leq i \leq n} c_{n,l,i}$, entonces,

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{2L_n^2 nc_n^2}} = 2de^{-\frac{(\epsilon - \gamma_n)^2}{2L_n^2 nc_n^2}}$$

Lema 3.8.9. Si para toda función $g_{n,l}$ $1 \leq l \leq d$ se cumple la condición dada por la ecuación 3.33, y además, $\frac{E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0$. Sea λ_n una sucesión positiva que tiende a 0, tal que para $0 < \alpha < 1/2$ $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\lambda_n)$ entonces para todo $\theta \in [0, 1]^d$:

$$(3.21) \quad \| G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \| < \lambda_n$$

Demostración: Sea $\Theta = (\Theta_1, \dots, \Theta_d)$ y $\theta = (\theta_1, \dots, \theta_d)$. Tenemos que:

$$G_{g_n}(E(g_n(X)|\Theta = \theta)) = (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta = \theta)), \dots, P(g_{n,d} \leq E(g_{n,d}(X)|\Theta = \theta))) = \\ = (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta_1 = \theta_1)), \dots, P(g_{n,d} \leq E(g_{n,d}(X)|\Theta_d = \theta_d)))$$

$$\text{Luego, } \| G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \| = \max_{1 \leq l \leq d} |P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l|$$

Trabajando con la componente l -ésima y considerando que las marginales son uniformes:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) = \int_0^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* = \\ = \int_0^{\theta_l + \frac{\lambda_n}{2}} P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* + \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq$$

$$E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* \leq \theta_l + \frac{\lambda_n}{2} + \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^*$$

Tenemos que: $\theta_l^* - \theta_l > \frac{\lambda_n}{2}$ y $E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) < 0$ pues $E(g_{n,l}(X)|\Theta_l)$ es creciente.

Además, $E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) = [\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l}](c)(\theta_l^* - \theta_l) > M_l \frac{\lambda_n}{2}$ con $c \in (\theta_l, \theta_l^*)$, de donde

$$g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \text{ si y solo si}$$

$$|g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*)| \geq E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) > M_l \frac{\lambda_n}{2}$$

$$\text{Luego, } \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l}(X) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* =$$

$$= \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \mid \Theta_l = \theta_l^*) d\theta_l^* <$$

$$< \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(|g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*)| > M_l \frac{\lambda_n}{2} \mid \Theta_l = \theta_l^*) d\theta_l^* \leq 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

De esto obtenemos que:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) < \theta_l + \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}}$$

para todo l con $1 \leq l \leq d$. Así pues, como $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\lambda_n)$ con $0 < \alpha < 1/2$, existe K , tal que $\frac{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha}{\lambda_n} < K$, entonces para todo l , tenemos $\lambda_n > \frac{1}{K} (\sum_{i=1}^n c_{n,l,i}^2)^\alpha$, de donde,

$$\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2} > \frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}$$

A partir de esto, $e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}$ y luego

$$\frac{e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}}}{\lambda_n} < K \frac{e^{-\frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}}{(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0, \text{ ya que } 2\alpha - 1 < 0 \text{ y } \sum_{i=1}^n c_{n,l,i}^2 \rightarrow 0.$$

Así para n suficientemente grande: $2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < \frac{\lambda_n}{2}$

Entonces para $n \geq n_0$: $P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \lambda_n$.

Teorema 3.8.10. Supongamos que se cumplen las hipótesis del lema 3.8.8 y del lema 3.8.9 para $0 < \alpha < 1/2$ y $\frac{\gamma_n}{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$

1) Si $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$

$\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad.

2) Si además $\lim \log n \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$, entonces, $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente (en particular con probabilidad 1).

Demostración: Consideremos una sucesión $\epsilon_n \rightarrow 0$ positiva, tal que $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$. (siendo K una de las cotas).

$$\begin{aligned} P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) &= P(\bigcup_{k=1}^N \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq \sum_{k=1}^N P(\|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) = \sum_{k=1}^N P(\|\hat{G}_N(g_n(X_k)) - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq \sum_{k=1}^N P(\|\hat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k))\| > \frac{\epsilon_n}{3}) + \\ &+ \sum_{k=1}^N P(\|G_{g_n}(g_n(X_k)) - G_{g_n}(E(g_n(X_k)))\| > \frac{\epsilon_n}{3}) + \sum_{k=1}^N P(\|G_{g_n}(E(g_n(X_k))) - \Theta_{n,k}\| > \frac{\epsilon_n}{3}) \end{aligned}$$

Ahora:

$$\begin{aligned} P(\|\hat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k))\| > \frac{\epsilon_n}{3}) &= P(\max_{l=1,\dots,d} |\hat{G}_{N,l}(g_n(X_k)) - G_{g_n,l}(g_n(X_k))| > \frac{\epsilon_n}{3}) \leq \\ &\leq \sum_{l=1}^d P(|\hat{G}_{N,l}(g_n(X_k)) - G_{g_n,l}(g_n(X_k))| > \frac{\epsilon_n}{3}) = \end{aligned}$$

$$\sum_{l=1}^d P(|\hat{F}_{N,l}(g_n(X_k)) - P(g_{n,l}(X) \leq g_{n,l}(X_k))| > \frac{\epsilon_n}{3}) \leq dCe^{-2\frac{\epsilon_n^2 N}{9}}$$

Para el segundo término, aplicamos el lema 3.8.8.

$$P(\|G_{n,k}(g_n(X)) - G_{n,k}(E(g_n(X)))\| > \frac{\epsilon_n}{3}) \leq 2 \sum_{l=1}^d e^{-\frac{(\frac{\epsilon_n}{3} - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$$

El último término, se anula para $n \geq n_0$. Así hemos obtenido que:

$$(3.22) \quad P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq NdCe^{-2\frac{\epsilon_n^2 N}{9}} + 2N \sum_{l=1}^d e^{-\frac{(\frac{\epsilon_n}{3} - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}} \text{ o sea,}$$

$$(3.23) \quad P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq Nd \left(Ce^{-2\frac{\epsilon_n^2 N}{9}} + 2e^{-\frac{\epsilon_n^2 (1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2 \max_l \sum_{i=1}^n c_{n,l,i}^2}} \right).$$

Si se cumple cualquiera de las dos condiciones, entonces: $e^{-2\frac{\epsilon_n^2 N}{9}} < e^{-\frac{2 \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{9K^2}}$, ya que para n suficientemente grande $3\frac{\gamma_n}{\epsilon_n} < 1/2$ para el segundo término tenemos:

$$e^{-\frac{\epsilon_n^2 (1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2 \max_l \sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{144K^2L_n^2}}. \text{ Así pues,}$$

$$P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) < N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

En el caso 1: $P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) < C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} (\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}$

o sea, $P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \rightarrow 0$, lo cual implica la convergencia en probabilidad.

Si se cumple 2, obtenemos que:

$$\lim_{n \rightarrow \infty} \frac{\frac{32L_n^2+1}{144K^2L_n^2} (\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}{\log n} = +\infty$$

lo que implica la convergencia completa pues si tenemos una serie: $\sum_{n=1}^{\infty} e^{-f(n)}$ y

$Q(n) = \frac{f(n)}{\log n} \rightarrow +\infty$ entonces la serie converge, pues, $e^{-f(n)} = e^{\log n^{-Q(n)}} = n^{-Q(n)}$ y si $Q(n) \rightarrow +\infty$ la $\sum_{n=1}^{\infty} n^{-Q(n)}$ es convergente.

Las series exponenciales convergen y la convergencia con probabilidad 1 se deduce de aplicar el lema de Borel Cantelli.

Observación: Si consideramos el caso unidimensional con $g_n(X) = \bar{X}$ y $nP(\sum_{i=1}^n X_i = k) < L$ vemos que: $\gamma_n = \frac{1}{n}$ y $c_n = \frac{L}{n}$, por lo que la condición pedida se cumple ya que $\frac{1/n}{1/n^\alpha} \rightarrow 0$.

Si consideramos ϵ_n tal que $\frac{1}{n} = O(\epsilon_n)$ entonces si $n < N < e^{\frac{32L^2+1}{144K^2L^2} n^{\frac{1}{2}-\alpha}}$ y dado que $\frac{\log n}{n^{1-2\alpha}} \rightarrow 0$ entonces, $\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente.

Por la forma de los estimadores se pueden producir empates; romperemos los empates sumando a $g_{n,l}$ una variable aleatoria uniforme.

Consideremos una sucesión a_n por ejemplo $a_n = \min_l \min_{X, X'} |g_{n,l}(X) - g_{n,l}(X')|$ con X y X' posibles vectores de respuestas. La sucesión debe cumplir $a_n < \max_l (\sum_{i=1}^n c_{n,l,i}^2)$

Sea: $Z_n = g_n(X) + W$ con W una variable aleatoria con distribución uniforme $U[0, a(n)]^d$ y consideremos las empíricas:

$$\hat{J}_{N,l}(x) = \frac{\#\{Z_{n,l} \leq x\}}{N} = \frac{\sum_{i=1}^N \chi_{\{Z_{n,l} \leq x\}}}{N}$$

con $l = 1, \dots, d$. Definamos la función: $\hat{T}_N(x_1, \dots, x_d) = (\hat{J}_{N,1}(x_1), \dots, \hat{J}_{N,d}(x_d))$

Entonces:

$$(3.24) \quad \hat{\Theta}_n^* = \hat{T}_N(Z_n)$$

Lema 3.8.11. Supongamos que para $0 < \alpha < 1/2$ $\frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$

Si $\epsilon_n \rightarrow 0$ es una sucesión positiva, tal que: $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$ (siendo K una de las cotas). Entonces existe una constante C_1 independiente de n , tal que:

$$P(\|\hat{\Theta}_n - \hat{\Theta}_n^*\| > \epsilon_n) \leq dC_1 e^{-\frac{2\epsilon_n^2 N}{9}}$$

Demostración:

$$P(\|\hat{\Theta}_n - \hat{\Theta}_n^*\| > \epsilon_n) = P(\|\hat{G}_N(g_n(X)) - \hat{T}_N(Z_n)\| > \epsilon_n) = P(\max_{1 \leq l \leq d} |\hat{F}_{N,l}(g_n(X)) - \hat{J}_{N,l}(Z_n)| > \epsilon_n).$$

Trabajemos con cada componente.

$$\begin{aligned} \hat{F}_{N,l}(x - a_n) &= \frac{\#\{g_{n,l}(X) \leq x - a_n\}}{N} \leq \frac{\#\{Z_{n,l} \leq x\}}{N} = \hat{J}_{N,l}(x) \leq \hat{J}_{N,l}(x + a_n) = \\ &= \frac{\#\{Z_{n,l} \leq x + a_n\}}{N} \leq \frac{\#\{g_{n,l}(X) \leq x + a_n\}}{N} = \hat{F}_{N,l}(x + a_n) \end{aligned}$$

Luego, $\hat{F}_{N,l}(g_{n,l}(X) - a_n) \leq \hat{J}_{N,l}(Z_n) \leq \hat{F}_{N,l}(g_{n,l}(X) + a_n)$. De aquí,

$$|\hat{J}_{N,l}(Z_n) - \hat{F}_{N,l}(g_{n,l}(X))| \leq \hat{F}_{N,l}(g_{n,l}(X) + a_n) - \hat{F}_{N,l}(g_{n,l}(X) - a_n)$$

A partir de esto,

$$\begin{aligned} P(|\hat{J}_{N,l}(Z_n) - \hat{F}_{N,l}(g_{n,l}(X))| > \epsilon_n) &\leq P(\hat{F}_{N,l}(g_{n,l}(X) + a_n) - \hat{F}_{N,l}(g_{n,l}(X) - a_n) > \epsilon_n) \leq \\ &\leq P(|\hat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n)| > \frac{\epsilon_n}{3}) + P(|F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon_n}{3}) + \end{aligned}$$

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon_n}{3})$$

Sabemos que:

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n)| > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}} y$$

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}}$$

$$F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n) \leq L_n(g_{n,l}(X) + a_n - g_{n,l}(X) - a_n) + \gamma_n = 2a_n L_n + \gamma_n$$

Pero para $n > n_0$ tenemos que: $\gamma_n < \frac{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha}{6K} < \frac{\epsilon_n}{6}$ y $a(n) < \frac{\epsilon_n}{12L_n}$, luego para $n > n_0$, el último término se anula, así pues tenemos que existe C_1 tal que para todo n :

$$P(\|\widehat{\Theta}_n - \widehat{\Theta}_n^*\| > \epsilon_n) \leq dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}.$$

Obtenemos como corolario que:

Corolario 3.8.12. Supongamos que para $0 < \alpha < 1/2$ $\frac{\gamma_n}{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$.

1) Si $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$ en probabilidad

$$\max_{k=1,\dots,N} \|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \rightarrow 0$$

2) Si además $\lim \log n \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$, entonces

$$\max_{k=1,\dots,N} \|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \rightarrow 0$$

completamente (en particular con probabilidad 1).

Demostración: Consideremos una sucesión $\epsilon_n \rightarrow 0$ positiva, tal que: $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$ (siendo K una de las cotas).

$$\|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \leq \|\widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k}\| + \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| \text{ luego}$$

$$\begin{aligned} P(\|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| > \epsilon_n) &\leq P(\|\widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k}\| > \epsilon_n) + P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq 2dCe^{\frac{-2\epsilon_n^2 N}{9}} + P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \end{aligned}$$

Aplicamos el teorema 3.8.10 a los términos $dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}$ y $P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n)$ de donde obtenemos el resultado.

Lema 3.8.13. Consideremos que se cumplen las hipótesis del teorema 3.8.10 y que el núcleo K es simétrico acotado y con soporte compacto C_K . Supongamos que las CCI son Lipchitz de constante M_V en un conjunto V compacto con $V \subset [0, 1]^d$. Además, para $0 < \alpha < 1/2$

$$N > \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)} \text{ y } \max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$$

Existe un entero n_0 tal que para todo $n > n_0$ y siendo $r = \max\{\|x\| \mid x \in C_K\}$ la siguiente desigualdad ocurre para todo ítem i y $\theta_0 \in U \subset V$, U abierto.

(3.25)

$$P(|\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq e^{-8 \frac{m_l^2 M_V^2 N n T h_n^{d+2}}{9 \|K\|_\infty^2 L^2 r^{d-2}}} + N_n(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}})$$

Demostración: Como $\frac{\sum_{k=1}^{N_n} K(\frac{\theta - \hat{\theta}_{n,i,k}}{h_n})}{N_n h_n^d}$ tiene límite finito con probabilidad 1. Existen m_l y m_u tal que para n suficientemente grande y $\theta \in (0, 1)^d$:

$$(3.26) \quad m_l N_n h_n^d < \sum_{k=1}^{N_n} K(\frac{\theta - \hat{\theta}_{n,i,k}}{h_n}) < m_u N_n h_n^d$$

en un conjunto de probabilidad 1.

Sabemos que $K(\frac{\theta - \hat{\theta}}{h_n}) > 0$ si y solo si $\frac{\theta - \hat{\theta}}{h_n} \in C_K$. Como C_K es compacto $C_K \subseteq B_K = \overline{B(0, r)}$ con $r = \max\{\|x\| \mid x \in C_K\}$

Sea: $A_{n,i,\theta} = \{k : \frac{\theta - \hat{\theta}_{n,i,k}}{h_n} \in B_K\} = \{k : \hat{\theta}_{n,i,k} \in \theta - h_n B_K\} = \{k : \hat{\theta}_{n,i,k} \in \overline{B(\theta, h_n r)}\}$ entonces,

$$(3.27) \quad \#A_{n,i,\theta} \rightarrow N_n \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})$$

donde, $\mu_{\hat{\theta}}$ es la distribución de $\hat{\theta}$, así pues en todo abierto U de $(0, 1)^d$ existe n_0 tal que para todo $\theta \in U$ y $n > n_0$, $|\#A_{n,i,\theta} - N_n \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})| < \epsilon$ de donde

$$(3.28) \quad \frac{N_n}{2} \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)}) < \#A_{n,i,\theta} < \frac{3N_n}{2} \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})$$

Sea $\theta_0 \in U$ y el suceso $B_{n,i} = \{\hat{\theta}_j - \theta_j \in h_n B_K \text{ para } j = 1, \dots, N_n\}$

$$\begin{aligned} P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) &= P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) P(B_{n,i}) + \\ &+ P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}^c) P(B_{n,i}^c) \end{aligned}$$

luego, $P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) + P(B_{n,i}^c)$

Como $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$ sabemos que:

$$P(B_{n,i}^c) = P(\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > h_n r) < N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Tenemos, entonces: $P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq$

$$\leq P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) + N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Sea $\Theta_n = (\Theta_{n,1}, \dots, \Theta_{n,N_n})$ y análogamente $\hat{\Theta}_{n,i}$ y sean θ_n y $\hat{\theta}_{n,i}$ realizaciones de estos vectores. Sea W la distribución conjunta de estos vectores condicionado a $B_{n,i}$ Tenemos que:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) = \int P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) dW(\hat{\theta}_{n,i}, \theta_n)$$

$$\begin{aligned} \text{Ahora: } P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) &\leq \\ &\leq P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| + |E(\hat{P}_{n,i}(\theta_0)) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) \end{aligned}$$

Esta esperanza condicional la podemos calcular como:

$$\begin{aligned} E(\hat{P}_{n,i}(\theta_0) | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) &= \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})} \sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n}) E(X_{n,i,k} | \theta_{n,j}) = \\ &= \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n}) P_{n,j}(\theta_{n,k}) \end{aligned}$$

Para todo $k \in A_{n,i,\theta_0}$ $\|\theta_0 - \theta_{n,k}\| \leq \|\theta_0 - \hat{\theta}_{n,-i,k}\| + \|\hat{\theta}_{n,-i,k} - \theta_{n,k}\| \leq h_n r + h_n r = 2h_n r$

Eligiendo n suficientemente grande para que $2h_nr < \inf\{\|x - y\| \mid x \in \partial U, y \in \partial V\}$ entonces que $\|\theta_0 - \theta_{n,k}\| \leq 2h_nr$ implica que para todo $k \in A_{n,i,\theta_0}$, $\theta_{n,k} \in V$. Luego, por ser las CCI Lipchitz en V :

$$|P_{n,i}(\theta_0) - P_{n,j}(\theta_{n,k})| \leq M_V \|\theta_0 - \theta_{n,k}\| \leq 2M_V h_n r$$

Esto implica que: $|E(\hat{P}_{n,i}(\theta_0)|B_{n,i}, \hat{\theta}_{n,i}, \theta_n) - P_{n,i}(\theta_0)| =$

$$= \left| \frac{1}{\sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n}) [P_{n,j}(\theta_{n,k}) - P_{n,i}(\theta_0)] \right| < 2M_V h_n r$$

Entonces para n suficiente grande:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) \leq P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n)$$

Una expresión alternativa para $\hat{P}_{n,i}(\theta_0)$ viene dada por:

$$\hat{P}_{n,i}(\theta_0) = \frac{1}{\#A_{n,i,\theta_0}} \sum_{j \in A_{n,i,\theta_0}} w_j X_{n,i,k} \text{ con } w_j = \frac{\#A_{n,i,\theta_0} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})}$$

A partir de 3.26 y 3.28 y como K es acotado para n suficientemente grande se tiene:

$$w_j \leq \frac{\frac{3N_n \|K\|_\infty}{2} \mu_{\hat{\theta}}(B(\theta, h_n r))}{m_l N_n h_n^d} \leq \frac{3r^d h_n^d \|K\|_\infty L}{2m_l h_n^d} = \frac{3r^d \|K\|_\infty L}{2m_l}$$

donde aplicamos el resultado debido a Devroye que $\frac{\mu_{\hat{\theta}}(B(\theta, h_n r))}{h_n^d r^d}$ tiene límite finito para casi todo θ con respecto a μ cuando h tiende a 0. Por lo tanto,

$$Th_n^d r^d < \mu_{\hat{\theta}}(B(\theta, h_n r)) < Lh_n^d r^d$$

Aplicando Hoeffding se obtiene:

$$\begin{aligned} & P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) = \\ & P\left(\left| \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k Y_{n,i,k} - \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k E(X_{n,i,k} | \theta_{n,k}) \right| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n\right) = \\ & \leq P\left(\left| \sum_{k \in A_{n,i,\theta_0}} [X_{n,i,k} - E(X_{n,i,k} | \theta_{n,k})] \right| > \#A_{n,i,\theta_0} \frac{2m_l M_V h_n}{3\|K\|_\infty L r^{d-1}} | B_{n,i}, \hat{\theta}_{n,i}, \theta_n\right) \\ & \leq 2e^{-\frac{8\#A_{n,i,\theta_0} m_l^2 M_V^2 h_n^2}{9\|K\|_\infty^2 L^2 r^{2(d-1)}}} \leq 2e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} \end{aligned}$$

Así pues, para n suficientemente grande y para todo $\theta_0 \in U$:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32L_n^2 + 1}{144K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Teorema 3.8.14. Dadas las hipótesis anteriores y para $\theta_0 \in (0, 1)^d$ fijo y

$$N > \eta \frac{9\|K\|_\infty^2 L^2 r^{d-2}}{m_l^2 M_V^2 T} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)}$$

$$(3.29) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

Observación: Si suponemos $N > \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}}$, entonces,

$$(3.30) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha})$$

Demostración: $ECM(\hat{P}_{n,i}(\theta_0)) = E((\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2) =$

$$= \int_0^1 P[(\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > t] dt \leq (3M_V h_n r)^2 + P[(\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > (3M_V h_n r)^2]$$

Esto es, $ECM(\hat{P}_{n,i}(\theta_0)) \leq (3M_{c,d} C_k h_n)^2 + P[|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r]$.

Aplicando el lema anterior tenemos para n suficientemente grande y todo ítem i :

$$ECM(\hat{P}_{n,i}(\theta_0)) \leq (3M_V h_n r)^2 + e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \|K\|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32L_n^2+1}{144K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right) \\ \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

En el caso unidimensional como $0 < \alpha < 1/2$ si se elige α próximo a $1/2$ entonces el ECM se hace cercano a $O(\frac{1}{n})$. Tenemos entonces el siguiente resultado:

Teorema 3.8.15. Dadas las hipótesis anteriores, suponiendo además que el núcleo es Lipchitz de constante J y, además, $N < Mn^r$ con $r > d + 1$

$$\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$$

si $n \rightarrow +\infty$ completamente con $(a, b)^d \subset (0, 1)^d$.

Demostración: Consideremos una grilla de $2^d N^{2d}$ puntos tal que la distancia (con la norma del máximo) entre dos puntos sea menor o igual a $1/N^2$. Sea θ un punto arbitrario de $[0, 1]^d$ entonces existe k tal que la distancia de θ_k^n a θ no es mayor a $1/N^2$.

Para todo θ tenemos

$$|P_{n,i}(\theta) - \hat{P}_{n,i}(\theta)| \leq |P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| + |P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| + |\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)|$$

Consideremos

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| = \left| \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} \right| \leq \left| \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} \right| + \\ + \left| \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \right| = \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \left| \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) - \sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) \right| + \\ + \frac{1}{\left| \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) \right|} \left| \sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij} - \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij} \right| \leq \frac{2 \sum_{i=1}^N |K(\frac{\theta_k^n - \hat{\theta}_i}{h}) - K(\frac{\theta - \hat{\theta}_i}{h})|}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \leq \frac{2 \sum_{i=1}^N J \|\theta_k^n - \theta\|}{h \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})}$$

Como $\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) > Nh^d m_l$ obtenemos que:

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| < \frac{2JN\|\theta_k^n - \theta\|}{Nh^{d+1}m_l} < \frac{2J}{N^2 h^{d+1} m_l}$$

Como existe C_1 tal que: $h > C_1 \max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha$ y $N > \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)}$

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| = o(\frac{1}{n})$$

A partir de $\sup_{\theta \in [0,1]^d} \left| \frac{\partial P_i}{\partial \theta_j} \right| < c_j$ para todo j entonces si $c = \max_j c_j$ con $j=1, \dots, d$

$$|P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| \leq c \|\theta - \theta_k^n\| \leq c \frac{1}{N^2}$$

Sea $\epsilon > 0$ mostraremos que: $\sum_{n=1}^\infty P(\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| > \epsilon) < \infty$

Para probar esto es suficiente probar que:

$$(3.31) \quad \sum_{n=1}^\infty \sum_{i=1}^n \sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > \epsilon/3] < \infty$$

ya que para $n > n_0$ suficientemente grande, $|P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| \leq \frac{\epsilon}{3}$ y $|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| \leq \frac{\epsilon}{3}$, entonces,

$$P\left(\sum_{n=1}^{\infty} P(\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| > \epsilon) < \sum_{n=1}^{\infty} \sum_{i=1}^n \sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > \epsilon/3]$$

Para este n_0 la triple suma es menor que:

$$(3.32) \quad 2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} \sum_{i=1}^n \sum_{k=1}^{2N_n^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > 3M_V h_n r]$$

Se puede aplicar el lema 3.8.13 y entonces 3.32 es menor que:

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} 2^d n N_n^{2d} \left(e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \|K\|_{\infty}^2 L^2 r^{d-2}}} + N_n (C_1 e^{-\frac{32 L_n^2 + 1}{144 K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}) \right)$$

Reemplazando N_n por su cota superior $M n^r$ entonces lo anterior queda acotado por

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} 2^d M^{2d} n^{2dr+1} \left(e^{-8 \frac{m_l^2 M_V^2 T \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - [d(1/2-\alpha)+1-2\alpha]}{9 \|K\|_{\infty}^2 L^2 r^{d-2}}} + \right. \\ \left. + M n^r (C_1 e^{-\frac{32 L_n^2 + 1}{144 K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}) \right)$$

Como $0 < \alpha < 1/2$ y $\lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$ la suma anterior es finita.

Ahora considerando en conjunto los resultados anteriores

Teorema 3.8.16. Si se cumplen las hipótesis:

- 1) Independencia Local
- 2) El rasgo Θ tiene distribuciones marginales uniformes $U[0, 1]$
- 3) para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$(3.33) \quad \sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2$$

para $1 \leq l \leq d$.

- 4) Las $F_{n,l}$ correspondientes son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$

$$5) \quad \frac{\partial E(g_{n,l}(X) | \Theta_l)}{\partial \Theta_l} > M_l > 0.$$

$$6) \quad \text{Para } 0 < \alpha < 1/2 \quad \frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha}} \rightarrow 0$$

$$7) \quad H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)} < N < M n^r \text{ con } r > d + 1$$

$$8) \quad \lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0.$$

- 9) El núcleo K es simétrico, acotado y con soporte compacto.

10) Las CCI tienen derivadas parciales continuas en $(0, 1)^d$ y en cada compacto $[a, b]^d \subset (0, 1)^d$ son acotadas

$$11) \quad \text{Para } 0 < \alpha < 1/2 \quad \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha} = O(h_n)$$

- 12) El núcleo K es Lipschitz de constante J .

entonces $\max\{\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1

Demostración:

Sale directamente de los teoremas 3.6.1 y 3.6.3 notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

3.9. Propiedades del estimador de la CCI si la habilidad es medida sin error

Si bien hemos demostrado la consistencia conjunta de los estimadores, en esta sección consideramos el caso más sencillo que es cuando los rasgos son conocidos o cuando son medidos sin error o sea $\widehat{\Theta}_i = \Theta_i$. Con esta simplificación podremos encontrar la distribución asintótica de las CCI. Ramsay (1991) aproxima la distribución para el caso de rasgo medido sin error pero con diseño fijo. En nuestro caso tomaremos en cuenta la distribución del rasgo en la población. El estimador en este caso es:

$$(3.34) \quad \widehat{P_j(\theta)} = \frac{\sum_{i=1}^N X_{i,j} K_H(\theta - \widehat{\Theta}_i)}{\sum_{i=1}^N K_H(\theta - \widehat{\Theta}_i)} = \frac{\sum_{i=1}^N X_{i,j} K_H(\theta - \Theta_i)}{\sum_{i=1}^N K_H(\theta - \Theta_i)}$$

Primeramente veamos que en este caso el estimador es consistente con hipótesis más simples y con un núcleo no necesariamente con soporte compacto. Lo que veamos servirá también para respuestas continuas cuando queremos calcular $E(X|\Theta)$

Teorema 3.9.1. Supongamos que $\Theta \sim f_\theta$, $f_\theta(\theta) > 0$ y continua en θ además la habilidad es medida sin error entonces si $h_N \rightarrow 0$, $Nh_N \rightarrow \infty$

$\int K(u)du = 1$ y $|\theta K(\theta)| \rightarrow 0$ cuando $|\theta| \rightarrow \infty$, $Nh^2 \rightarrow \infty$ y K acotada entonces el estimador de la CCI 3.34 es consistente.

Demostración:

$$\widehat{P}(\theta) = \frac{\frac{1}{Nh} \sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right)}{\frac{1}{Nh} \sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right)} = \frac{R_{1N}(\theta)}{R_{0N}(\theta)}$$

Bajo estas condiciones se tiene que, $R_{0N}(\theta) \rightarrow f(\theta)$

Trabajemos ahora con el término $R_{1N}(\theta)$,

$$E(R_{1N}(\theta)) = E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) X_i\right) = \frac{1}{h} E\left(K\left(\frac{\theta - \Theta_i}{h}\right) E(X_i|\Theta)\right) = \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) P(t) f_\Theta(t) dt \rightarrow P(\theta) f(\theta)$$

en cada punto de continuidad de Pf por el lema 3.3.1

$P.f \in L^1$, pues $\int |P(t)| f(t) dt = E(|P(\theta)|) = E(|E(X/\Theta)|) \leq E(E(|X|/\Theta)) = E(|X|)$ o sea basta que $E(|X|) < \infty$ lo cual claramente se cumple (por el tipo de variable que consideramos) y las hipótesis sobre K para aplicar el lema 3.3.1.

Probemos ahora que la varianza tiende a cero:

$$\begin{aligned} Var(R_{1N}(\theta)) &= \frac{1}{Nh^2} Var\left(K\left(\frac{\theta - \Theta_1}{h}\right) X_1\right) \leq \frac{1}{Nh^2} E\left(X_1^2 K^2\left(\frac{\theta - \Theta_1}{h}\right)\right) = \\ &= \frac{1}{Nh^2} E\left(K^2\left(\frac{\theta - \Theta_1}{h}\right) E(X_1^2|\Theta_1)\right) = \frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt \text{ donde} \\ m_2(\Theta) &= E(X^2|\Theta) \text{ y } E(m_2(\Theta)) = E(E(X^2|\Theta)) = E(X^2) \end{aligned}$$

Analicemos ahora el término $\frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt$.

Como $E(X^2) < \infty$, $Nh^2 \rightarrow \infty$ y K está acotado tenemos,

$$\frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt \rightarrow 0$$

Veamos ahora dos resultados, uno de convergencia en probabilidad y otro de convergencia completa.

Sean $(\Theta_1, X_1) \dots (\Theta_N, X_N)$ vectores aleatorios independientes, idénticamente distribuidas, con $\Theta_i \in R^d, X_i \in R$. Consideremos las hipótesis:

$$H1) h_N \rightarrow 0 \text{ y } Nh_N^d \rightarrow \infty$$

Para el núcleo $K : R^d \rightarrow R^+$ pidamos:

H2) existen $C_1, C_2 > 0$ y H tales que $C_1 H(\|\Theta\|) \leq K(\theta) \leq C_2 H(\|\Theta\|)$ con H acotada, decreciente y tal que $t^d H(t) \rightarrow 0$ si $|t| \rightarrow \infty$

H3) Existen $c, r > 0$ tal que $K(\theta) \geq c \chi_{\|\theta\| \leq r}(\theta)$

Si f es integrable y Θ tiene distribución μ entonces definimos:

$$\mu_h = \int K\left(\frac{\Theta-t}{h}\right) f(t) \mu(dt) / \int K\left(\frac{\Theta-t}{h}\right) \mu(dt) = E\left(K\left(\frac{\Theta-\bar{\Theta}}{h}\right) f(\bar{\Theta})\right) / E\left(K\left(\frac{\Theta-\bar{\Theta}}{h}\right)\right)$$

Observaciones:

Lo que le pedimos al núcleo es menos que suponer continuidad y soporte compacto, o radial, comúnmente se toma $K(\|t\|)$. La segunda hipótesis sobre el núcleo implica menos que pedir que $K(0) \neq 0$ y K continua en 0, podemos ver que esta condición la cumplen los núcleos usualmente utilizados. También debemos observar que no hay hipótesis sobre la distribución de las Θ . Presentamos primero dos lemas, el primero un resultado de teoría de la medida y el otro un lema de convolución.

Lema 3.9.2. Sea S_r la bola de radio r centrada en θ , y sea $a_h(\theta) = \frac{h^d}{\mu(S_h)}$ entonces $a_h(\theta)$ tiene un límite finito ctp $\theta(\mu)$.

Demostración: Sea λ la medida de Lebesgue, descompongámosla en la parte absolutamente continua respecto a μ en la parte singular. Notemos por λ_1 a la parte absolutamente continua de λ respecto a μ y λ_2 la parte singular, luego

$$\frac{\lambda(S_h)}{\mu(S_h)} \longrightarrow \frac{d\lambda_1}{d\mu}(\theta) \text{ ctp } \theta(\mu) \text{ como } \lambda(S_h) = h^d \lambda(B_1)$$

$$\text{entonces } \frac{h^d}{\mu(S_h)} = \frac{\lambda(S_h)}{\mu(S_h)} \frac{1}{\lambda(B_1)} \longrightarrow \ell(\theta) \text{ ctp } \theta(\mu) \text{ donde } \ell \text{ es no negativa y finita.}$$

Lema 3.9.3. Sea K no negativo verificando H2) y H3) y f integrable, entonces,

$$\mu_h(\theta) \rightarrow f(\theta) \text{ c.t.p. } \theta(\mu) \text{ si } h \rightarrow 0$$

Si K es el indicador de la bola unitaria, tenemos el teorema de diferenciación respecto a la medida μ . Lo que presentamos es cómo mezclar un teorema de convolución con un teorema de diferenciación.

$$\begin{aligned} \text{Demostración: } |\mu_h(\theta) - f(\theta)| &= \left| \frac{\int K\left(\frac{\Theta-y}{h}\right) f(y) \mu(dy)}{\int K\left(\frac{\Theta-y}{h}\right) \mu(dy)} - f(\theta) \right| = \left| \frac{\int K\left(\frac{\Theta-y}{h}\right) (f(y) - f(\theta)) \mu(dy)}{\int K\left(\frac{\Theta-y}{h}\right) \mu(dy)} \right| \leq \\ &\leq \frac{C_2}{C_1} \int H\left(\frac{\|\theta-y\|}{h}\right) |f(y) - f(\theta)| \mu(dy) / \int H\left(\frac{\|\theta-y\|}{h}\right) \mu(dy) \text{ por H2)} \end{aligned}$$

$$H(t) = \int_0^{H(t)} ds = \int_0^{+\infty} I_{\{H(t) > s\}} ds \text{ y si } A_{t,h} = \left\{ y : H\left(\frac{\|\theta-y\|}{h}\right) > t \right\} \text{ entonces } H\left(\frac{\|\theta-y\|}{h}\right) = \int_0^{+\infty} I_{A_{t,h}}(t) dt$$

$$\text{Sea } \int H\left(\frac{\|\theta-y\|}{h}\right) \mu(dy) = \int_{-\infty}^{+\infty} \int_0^{+\infty} I_{A_{t,h}}(t) dt \mu(dy) = \int_0^{+\infty} \mu(A_{t,h}) dt \text{ ya que } I_{A_{t,h}} \text{ es integrable, aplico Fubini.}$$

$$\text{Ahora } \int H\left(\frac{\|\theta-y\|}{h}\right) |f(y) - f(\theta)| \mu(dy) = \int_{-\infty}^{+\infty} \int_0^{+\infty} I_{A_{t,h}}(t) |f(y) - f(\theta)| dt \mu(dy) = \int_0^{+\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt$$

aplicando nuevamente Fubini.

Dividiremos el cociente en dos partes, I y II. Dado $\epsilon > 0$, sea $\delta = \epsilon h^d$

$$I) \frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_0^{\infty} \mu(A_{t,h}) dt} \leq \frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_{\delta}^{\infty} \mu(A_{t,h}) dt} \leq \sup_{t \geq \delta} \frac{\int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\mu(A_{t,h})}$$

La última desigualdad es cierta ya que si consideramos dos funciones $g \geq 0$ y $h \geq 0$ con sus

integrales finitas entonces, $\frac{\int_{\delta}^{\infty} g dt}{\int_{\delta}^{\infty} h dt} \leq \sup_{t \geq \delta} \{g/h\}$ Supongamos que $\frac{P(t)}{h(t)} < A$ para todo $t \geq \delta$.

Como g y h son no negativas

$$\int_{\delta}^{+\infty} g dt \leq A \int_{\delta}^{+\infty} h(t) dt \text{ luego } \frac{\int_{\delta}^{+\infty} g(t) dt}{\int_{\delta}^{+\infty} h(t) dt} \leq A$$

Cualquier cota de g/h cumple la desigualdad, así pues lo cumple el supremo de g/h . Si $t \geq \delta$, consideremos los conjuntos $A_{t,h}$ y $A_{\delta,h}$, se tiene $radio(A_{t,h}) < radio(A_{\delta,h})$. Tenemos que $radio(A_{\delta,h}) \rightarrow 0$ pues: $\left\{y : H\left(\frac{\|\theta - y\|}{h}\right) > \delta\right\} = \{y : \|\theta - y\| < H^{-1}(\delta)h\}$ y $hH^{-1}(\delta) = hH^{-1}(\epsilon h^d) \rightarrow 0$ pues $t^d H(t) \xrightarrow{|t| \rightarrow \infty} 0$

Veamos lo anterior. Supongamos que $H^{-1}(\epsilon h^d) = t$, entonces se tiene que $h = \left(\frac{H(t)}{\epsilon}\right)^{1/d}$ luego $|hH^{-1}(\epsilon h^d)| = |t \left(\frac{H(t)}{\epsilon}\right)^{1/d}| \rightarrow 0$ si $h \rightarrow 0$ pues $t \rightarrow \infty$. Lo último se deduce ya que $t^d H(t) \rightarrow 0$ $|t| \rightarrow \infty$ y es claro que $t(H(t))^{1/d} \rightarrow 0$ si $|t| \rightarrow \infty$. Utilizamos un resultado de desigualdades para funciones maximales debido a Hardy y Littlewood. Concretamente, el resultado nos dice que si $radio(A_{t,h}) \rightarrow 0$ entonces

$$\sup_{t \geq \delta} \frac{\int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\mu(A_{t,h})} \rightarrow 0 \text{ ctp } \theta(\mu)$$

Veamos ahora la segunda parte

$$II) \frac{\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_0^{\infty} \mu(A_{t,h}) dt}$$

Llamemos $C = \int |f(y)| \mu(dy)$, entonces $\int_0^{\delta} \int_{\Delta_{t,h}} f(y) \mu(dy) \mu \leq C\delta$ y $\int_0^{\delta} \int_{A_{t,h}} f(\theta) \mu(dy) dt \leq \delta f(\theta)$

así pues $\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt \leq \delta(C + f(\theta))$

También $\int_0^{\infty} \mu(A_{t,h}) dt = \int H\left(\frac{\|\theta - y\|}{h}\right) \mu(dy) \geq C_1 \mu(S_{rh}) = \frac{C_1 h^d r^d}{a_{rh}(\theta)}$ luego

$$\frac{\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\int_0^{\infty} \mu(A_{t,h}) dt} \leq \frac{\delta(C + f(\theta)) a_{rh}(\theta)}{C_1 r^d h^d} = \frac{(f(\theta) + C) \epsilon a_{rh}(\theta)}{C_1 r^d} < \epsilon' \text{ ctp } \theta(\mu)$$

Aplicamos en la última desigualdad el lema 3.9.2

Teorema 3.9.4. Bajo H1), H2) y H3) entonces:

$$P_N(\theta) \xrightarrow{P} P(\theta) \text{ en ctp } \theta(\mu) \text{ donde } P_N(\theta) = \frac{\sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right)}{\sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right)}$$

Obs: En el caso que consideramos respuestas continuas y queremos estimar su esperanza condicional debemos pedir además que $E(|X|) < \infty$ condición necesaria para que exista la esperanza condicional.

Demostración: dividamos por $E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$ obteniendo:

$$P_N(\theta) = \frac{\sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)}{\sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right) / E\left(K\left(\frac{\theta - \Theta_j}{h}\right)\right)}$$

Notemos cómo:

$$V_{Ni} = X_i K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \quad Z_{Ni} = K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \quad B_{1N} = \frac{1}{N} \sum_{i=1}^N (V_{iN} - E(V_{iN}))$$

$$B_{2N} = \frac{1}{N} \sum_{i=1}^N (Z_{iN} - E(Z_{iN})) \quad A_N = E(V_{Ni}) = E\left(X_i K\left(\frac{\theta - \Theta_i}{h}\right)\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

Observamos que $E(Z_{Ni}) = 1$ y $P_N(\theta) = \frac{(B_{1N} + A_N)}{(B_{2N} + 1)}$. Ahora,

$$A_N = \frac{E\left(K\left(\frac{\theta - \Theta_i}{h}\right) E(X/\Theta)\right)}{E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)}$$

así pues, aplico el lema 3.9.3, tomando $f = P(\theta)$; tenemos $P(\theta) \in L^1$, por ser $E(|X|) < \infty$; $K \geq 0$ por H3 y $A_N = \mu_h(\theta)$, entonces, $A_N \rightarrow P(\theta)$ ctp $\theta(\mu)$

Entonces el problema se reduce a probar que $B_{1N} \xrightarrow{P} 0$ y $B_{2N} \xrightarrow{P} 0$ ctp $\theta(\mu)$

Probemos primero que $B_{1N} \xrightarrow{P} 0$. Sean $N > 0$ $Y' = X \chi_{\{|X| \leq N\}}$ e $Y'' = X - Y'$

$$B'_{1N} = \sum_{i=1}^N [Y'_i K\left(\frac{\theta - \Theta_i}{h}\right) - E(Y' K\left(\frac{\theta - \Theta_i}{h}\right))] / n E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

$$B''_{1N} = \sum_{i=1}^N [Y''_i K\left(\frac{\theta - \Theta_i}{h}\right) - E(Y'' K\left(\frac{\theta - \Theta_i}{h}\right))] / n E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

Se tiene por la linealidad de la esperanza que $B'_{1N} + B''_{1N} = B_{1N}$. Aplicando Chebichev

$$P(|B'_{1N}| > t) \leq \frac{1}{N} \text{Var}[Y'_i K\left(\frac{\theta - \Theta_i}{h}\right)] / t^2 E^2\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \leq \frac{1}{N t^2} \frac{E(Y'^2 K^2\left(\frac{\theta - \Theta_i}{h}\right))}{E^2(K\left(\frac{\theta - \Theta_i}{h}\right))} \leq \frac{N^2 \|K\|_\infty}{N t^2} \frac{1}{E(K\left(\frac{\theta - \Theta_i}{h}\right))}$$

Por H3: $E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \geq c E\left(\chi_{\{\|u\| \leq r\}}\left(\frac{\theta - \Theta}{h}\right)\right) = c \int_{\|\theta - t\| \leq rh} \mu(dt) = c \mu(S_{rh})$ luego

$$\frac{N^2 \|K\|_\infty}{Nt^2} \frac{1}{E(K(\frac{\theta-\Theta_i}{h}))} \leq \frac{N^2 \|K\|_\infty}{Nt^2} \frac{r^d h^d}{c\mu(S_{rh})r^d h^d} = \frac{N^2 \|K\|_\infty}{Nt^2 c r^d h^d} a_{rh} = \frac{cte}{N h^d} \rightarrow 0 \text{ ctp } \theta(\mu)$$

para cada N , pues a_{rh} tiene límite finito. Así pues para cada N fijo, $B'_{1N} \xrightarrow{P} 0$ ctp $\theta(\mu)$

Aplicando Markov a B''_{1N} y siendo $g_N(\theta) = E(|Y''|/\Theta)$

$$P(|B''_{1N}| > t) \leq \frac{E(|B''_{1N}|)}{t} \leq \frac{2}{t} \frac{E(|Y''|K(\frac{\theta-\Theta}{h}))}{E(K(\frac{\theta-\Theta}{h}))} = \frac{2}{t} \frac{E(K(\frac{\theta-\Theta}{h}) E(|Y''|/\theta))}{E(K(\frac{\theta-\Theta}{h}))} = \frac{2}{t} g_{n,N}(\theta)$$

Por el lema 3.9.3 con $f = P_N$ se tiene que $\frac{2}{t} g_{n,N}(\theta) \rightarrow P_N(\theta) \frac{2}{t}$ ctp $\theta(\mu)$,

$P_N(\theta) = E(|Y''|/\Theta) = E(|X - Y'|/\Theta), |X - Y'| < 2|X|$, y $|X - Y'| \rightarrow 0$ c.s. $N \rightarrow \infty$

Así pues, $E(P_N(\theta)) = E(E(|X - Y'|/\Theta)) = E(|X - Y'|) \rightarrow 0$ si $N \rightarrow \infty$, además, $P_N \geq 0$ y P_N decreciente, luego $P_N(\theta) \rightarrow 0$ ctp $\theta(\mu)$ con $N \rightarrow \infty$,

Elijamos N tal que $P_N(\theta)$ sea pequeño, entonces si $N \rightarrow \infty$ $B_{1N} \xrightarrow{P} 0$

Esto vale en $A_0 = \{\theta : P_N(\theta) \rightarrow 0\}$ y $A_N = \{\theta : g_{Nh} \rightarrow P_N\}$ y vale en $\bigcap_{N=0}^{\infty} A_N$

Para B_{2N} se hace lo mismo pero como no están las X no hay que truncar.

Teorema 3.9.5. Convergencia completa

Si $|X| < \gamma < \infty$, y se verifican H1), H2), H3), $\frac{Nh^d}{\log N} \rightarrow \infty$ entonces $g_N(\theta) \xrightarrow{c} P(\theta)$ ctp $\theta(\mu)$

Demostración: Escribimos $P_N(\theta) = \frac{(B_{1N} + A_N)}{(B_{2N} + 1)}$. Sabemos que $A_N \rightarrow P(\theta)$ ctp $\theta(\mu)$ por el teorema anterior, luego basta ver que $B_{1N} \xrightarrow{c} 0$ y $B_{2N} \xrightarrow{c} 0$ ctp $\theta(\mu)$. Aplicaremos la desigualdad de Bernstein a V_{Ni}

$$|V_{Ni}| \leq \frac{\gamma \|K\|_\infty}{E(K(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma \|K\|_\infty a_{rh}(\theta)}{c r^d h^d}$$

que como ya vimos tiene límite finito.

$$Var(V_{Ni}) \leq E(V_{Ni}^2) = \frac{E(X_i^2 K^2(\frac{\theta-\Theta}{h}))}{E^2(X_i^2 K^2(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma^2 \|K\|_\infty^2}{E(K(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma^2 \|K\|_\infty^2 a_{rh}(\theta)}{c r^d h^d} \text{ entonces}$$

$$\begin{aligned} P(|B_{1N}| > t) &\leq 2 \exp \left[\frac{-Nt^2}{2 \left(\frac{\gamma^2 \|K\|_\infty^2 a_{rh}(\theta)}{c r^d h^d} + \frac{\gamma \|K\|_\infty a_{rh}(\theta) t}{c r^d h^d} \right)} \right] = 2 \exp \left[\frac{-Nt^2 c r^d h^d}{2 \|K\|_\infty a_{rh}(\theta) \gamma (\gamma + t)} \right] \leq \\ &\leq 2 \exp \left(\frac{-C_{1N} h^d}{C_2 (\gamma + t)} \right) \text{ ctp } \theta(\mu). \text{ llamando a } \frac{c_1}{C_2 (\gamma + t)} = C \text{ obtenemos:} \end{aligned}$$

$$P(|B_{1N}| > t) \leq 2 \exp(-C(\theta) N h^d) \text{ y como } \frac{N h^d}{\log N} \rightarrow \infty \text{ entonces: } \sum 2 \exp(-C(\theta) N h^d)$$

converge, luego $\sum P(|B_{1N}| > t)$ converge, por lo que B_{1N} converge completamente.

Análogamente se procede con B_{2N} .

Veamos ahora los resultados para el caso general de los estimadores. Daremos sin demostración los resultados cuando las habilidades son conocidas.

Teorema 3.9.6. Si las habilidades $\Theta_1, \dots, \Theta_N$ son conocidos o medidas sin error se tiene para el estimador $\hat{P}(\theta)$ que

$$E(\hat{P}(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \mu_2(K) \frac{\nabla_P(\theta)^t H H^t \nabla_P(\theta)}{f(\theta)} + \frac{1}{2} \mu_2(K) \text{tr}(H^t W_P(\theta) H)$$

$$V(\hat{P}(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

Teorema 3.9.7. Si las habilidades $\Theta_1, \dots, \Theta_N$ son conocidos o medidas sin error se tiene para el estimador $\hat{P}^1(\theta)$ que

$$E(\hat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \frac{1}{2} \mu_2(K) \text{tr}(H^t W_P(\theta) H)$$

$$V(\hat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

3.9.1. Distribución asintótica de la CCI

Daremos un resultado para los rasgos conocidos o medidos sin error

Teorema 3.9.8. Sean $\{(\Theta_i, X_i) i \geq 1\}$ el rasgo y la respuesta a cada ítem y que son variables aleatorias independientes idénticamente distribuidas con $\Theta_i \in R^d$ y $X_i \in R$.

Notemos por $F(X_i/\Theta_i = u)$ a la distribución condicional de X_i , dado $\Theta_i = u$ y por $P(u) = E(X_1/\Theta_1 = u)$ y $\sigma^2(u) = E((X_1 - P(u))^2/\Theta_1 = u)$. Supongamos que:

- i) P es Lipchitz, y además $\lim_{\epsilon \rightarrow 0} \frac{P(\theta + \epsilon u) - P(\theta)}{\epsilon} = P'(\theta, u)$ es decir, P tiene derivada en la dirección de u (si P es diferenciable se cumplen las 2 condiciones)
- ii) $\sigma^2(\cdot)$ es continua en θ
- iii) $\Theta_i \sim f$ continua en θ y acotada
- iv) $K : R^d \rightarrow R$ acotado no negativo, $\int K = 1$; $|t|^d K(t) \rightarrow 0$ si $|t| \rightarrow \infty$ y $t^2 K^2(t)$ integrable
- v) existe $0 \leq \beta < \infty$ tal que $Nh^{d+2} \rightarrow \beta$

Sea $P_N(\theta) = \sum \omega_{N_i}(\theta) X_i$ con $\omega_{N_i}(\theta) = \frac{K \frac{\theta - \Theta_i}{h}}{\sum_{i=1}^N K \left(\frac{\theta - \Theta_i}{h} \right)}$, entonces,

$$(Nh^d)^{1/2} (P_N(\theta) - P(\theta)) \xrightarrow{W} N \left(b, \frac{\sigma^2(\theta)}{f(\theta)} \int K^2(u) du \right) \text{ con } b = \beta^{1/2} \int K(t) g'(\theta, t) dt$$

Observación: notando que $Nh^{d+2} \rightarrow \beta$ es $Nh^d h^2 \rightarrow \beta$ podemos observar que como $h^2 \rightarrow 0$ debe cumplirse que $Nh^d \rightarrow \infty$.

Demostración: Nuestro objetivo es calcular la distribución asintótica de $P_N(\theta) - P(\theta)$.

$$\begin{aligned} (Nh^d)^{1/2} \left[\frac{\sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i}{\sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} - P(\theta) \right] &= (Nh^d)^{1/2} \left[\frac{\sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i - P(\theta) \sum_j K \left(\frac{\theta - \Theta_j}{h} \right)}{\sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} \right] = \\ &= (Nh^d)^{1/2} \frac{\left[\frac{1}{Nh^d} \sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i - P(\theta) \frac{1}{Nh^d} \sum_j K \left(\frac{\theta - \Theta_j}{h} \right) \right]}{\frac{1}{Nh^d} \sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} \end{aligned}$$

Como $Nh^d \rightarrow \infty$ por la observación previa, $h \rightarrow 0$ y por iii), iv) y v) obtenemos

$$\frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \xrightarrow{P} f(\theta)$$

Analicemos ahora el término

$$\begin{aligned} & (Nh^d)^{1/2} \left[\frac{1}{Nh^d} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) X_i - P(\theta) \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \right] = \\ & \frac{1}{(Nh^d)^{1/2}} \left[\sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\theta)) \right] = \frac{1}{(Nh^d)^{1/2}} \left[\sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i) + P(\Theta_i) - P(\theta)) \right] = \\ & \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i)) + \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \end{aligned}$$

Notemos $V_{Ni} = K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i))$. Las variables V_{Ni} tienen media 0, y son independientes. Tenemos un sistema triangular y se verifican las condiciones de Lindeberg. Calculemos la varianza límite. Condicionando $(X_i - P(\theta))^2$ a Θ

$$\frac{1}{h^d} E(V_{Ni}^2) = \frac{1}{h^d} \int K^2\left(\frac{\theta - t}{h}\right) \sigma^2(t) f(t) dt \rightarrow \sigma^2(\theta) f(\theta) \int K^2$$

donde hemos aplicado el lema 3.3.1 ya que $\sigma^2 f$ es continua en θ por las hipótesis ii) y iii)

$$\text{Var} \left(\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} \right) = \frac{1}{h^d} E(V_{Ni}^2) \rightarrow \sigma^2 f(\theta) \int K^2(t) dt$$

Así pues, aplicando el teorema de Lindeberg tenemos que:

$$\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} \rightarrow N(0, \sigma^2 f \int K^2) \text{ luego,}$$

$$\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} / \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \rightarrow N\left(0, \frac{\sigma^2 f(\theta) \int K^2}{f(\theta)}\right)$$

$$\text{Veamos que: } \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \xrightarrow{P} \beta^{1/2} f(\theta) \int K(t) g'(\theta, t) dt$$

$$\begin{aligned} & E \left(\frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \right) = \frac{1}{(Nh^d)^{1/2}} N \frac{h^d}{h^d} \int K\left(\frac{\theta - t}{h}\right) (P(t) - P(\theta)) f(t) dt = \\ & = (Nh^d)^{1/2} \int K(u) (P(\theta - uh) - P(\theta)) f(\theta - uh) du = (Nh^d)^{1/2} \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) h du \end{aligned}$$

$$\text{Ahora si } h \rightarrow 0 \quad K(u) \frac{P(\theta - uh) - P(\theta)}{h} f(\theta - uh) \rightarrow K(u) P'(\theta, u) f(\theta)$$

Además f es acotada, $K(u) \frac{P(\theta - uh) - P(\theta)}{h} f(\theta - uh) \leq C |K(u)| |u|$ y esta es una función integrable, pues $u^2 K^2$ es integrable, luego puedo entonces aplicar convergencia dominada.

Nota: En vez de pedir f acotada puedo pedir K de soporte compacto, entonces fuera de un compacto $K(u)$ vale cero, por lo tanto allí puedo hacer uh suficientemente pequeño, de donde $f(\theta - uh) \leq |f(\theta) + \epsilon|$ y la acotación sería $C |k(u)| |u| |f(\theta) + \epsilon|$.

$$\text{Aplicando convergencia dominada: } \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) du \rightarrow f(\theta) \int K(t) P'(\theta, t) dt$$

$$\text{y } (Nh^d)^{1/2} h = (Nh^{d+2})^{1/2} \rightarrow \beta^{1/2} \text{ así pues,}$$

$$(Nh^d)^{1/2} h \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) du \rightarrow \beta^{1/2} f(\theta) \int K(t) P'(\theta, t) dt \text{ luego}$$

$$(Nh^d)^{1/2} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) (P(\theta - uh) - P(\theta)) / \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \rightarrow \beta^{1/2} \int K(t) P'(\theta, t) dt$$

Así pues tengo $\frac{a_N + b_N}{c_N}$ y $c_N \xrightarrow{P} c$; $a_N \xrightarrow{D} N$ y aplicando Slutsky $\frac{b_N}{c_N} \xrightarrow{P} \frac{b}{c}$

Veamos ahora que la varianza tiende a 0:

$$Var\left(\frac{1}{(Nh^d)^{1/2}} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))\right) = \frac{1}{(Nh^d)} NVar\left(K\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))\right) =$$

$$\frac{1}{h^d} E\left(K^2\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))^2\right) = \frac{1}{h^d} \int K^2\left(\frac{\theta - t}{h}\right) (P(t) - P(\theta))^2 f(t) dt =$$

$$\int K^2(u) (P(\theta - uh) - P(\theta))^2 f(\theta - uh) dt = h^2 \int K^2(u) \frac{(P(\theta - uh) - P(\theta))^2}{h^2} f(\theta - uh) du$$

ahora $\int K^2(u) \frac{(P(\theta - uh) - P(\theta))^2}{h^2} f(\theta - uh) du \rightarrow f(\theta) \int K^2(u) (g'(\theta, u))^2 du$ la integral del lado derecho existe pues g es Lipchitz y además $K^2 u^2$ es integrable, así pues aplico convergencia dominada y se tiene que si $h^2 \rightarrow 0$ la varianza $\rightarrow 0$.

3.10. Sesgo y varianza del estimador multidimensional

Siguiendo el mismo razonamiento que en el caso unidimensional tanto para el estimador no paramétrico multidimensional (basado en Nadaraya-Watson) como el estimador por polinomios locales de grado 1 la varianza condicional queda

$$V(P(\theta) | \widehat{\Theta_1, \dots, \Theta_N}) = \frac{P(\theta)Q(\theta)}{n|H|f(\theta)} \|K\|_2^2 (1 + o_p(1))$$

Para determinar el sesgo se aplica desarrollo de Taylor obteniéndose

$$(P(\Theta_1), \dots, P(\Theta_N))^t = P(\theta) Id_N + \begin{pmatrix} (\Theta_1 - \theta)^t \nabla_P(\theta) \\ \vdots \\ (\Theta_N - \theta)^t \nabla_P(\theta) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (\Theta_1 - \theta)^t \mathfrak{H}_P(\theta) (\Theta_1 - \theta) \\ \vdots \\ (\Theta_N - \theta)^t \mathfrak{H}_P(\theta) (\Theta_N - \theta) \end{pmatrix}$$

o lo que es igual

$$(P(\Theta_1), \dots, P(\Theta_N))^t = \begin{pmatrix} 1 & (\Theta_1 - \theta)^t \\ \vdots & \vdots \\ 1 & (\Theta_N - \theta)^t \end{pmatrix} \begin{pmatrix} P(\theta) \\ \nabla_P(\theta) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (\Theta_1 - \theta)^t \mathfrak{H}_P(\theta) (\Theta_1 - \theta) \\ \vdots \\ (\Theta_N - \theta)^t \mathfrak{H}_P(\theta) (\Theta_N - \theta) \end{pmatrix}$$

Reemplazando cada término como en dimensión 1 obtenemos:

$$E(\widehat{P(\theta)}) \approx \frac{1}{f(\theta) + o_p(1)} \left(\frac{\sum_{i=1}^N K_H(\Theta_i - \theta) P(\theta)}{n} + \sum_{i=1}^N K_H(\Theta_i - \theta) (\Theta_i - \theta)^t \nabla_P(\theta) \right. \\ \left. + \sum_{i=1}^N K_H(\Theta_i - \theta) (\Theta_i - \theta)^t \mathfrak{H}_P(\theta) (\Theta_i - \theta) \right)$$

o sea

$$E(\widehat{P(\theta)}) \approx \frac{1}{f(\theta)} \left(f(\theta)P(\theta) + \mu_2(K)\nabla_P H H^t \nabla_f + \frac{1}{2}\mu_2(K)f(\theta)\text{tr}(H^t \mathfrak{H}_P(\theta)H) \right)$$

esto último dado que

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta) = (f(\theta) + o_p(1))$$

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta)(\Theta_i - \theta) = \mu_2(K)HH^t \nabla_f + o_p(HH^t Id_d)$$

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta)(\Theta_i - \theta)(\Theta_i - \theta)^t = \mu_2(K)f(\theta)HH^t \nabla_f + o_p(HH^t)$$

por lo tanto el sesgo queda aproximadamente igual a $\frac{\mu_2(K)\nabla_P H H^t \nabla_f}{f(\theta)} + \frac{1}{2}\mu_2(K)\text{tr}(H^t \mathfrak{H}_P(\theta)H)$

En el caso del polinomio local de grado 1 operando de igual manera sobre la fórmula que utiliza e_1 se tiene que $e_1^t(P(\theta), \nabla_P(\theta))^t = P(\theta)$ luego el sesgo esta constituido por el segundo sumando o sea el sesgo es

$$\frac{1}{2}\mu_2(K)\text{tr}(H^t \mathfrak{H}_P(\theta)H)$$

3.10.1. Métodos de selección de ventanas

Veremos el método de validación cruzada. Recordemos que la suma de los cuadrados residuales es:

$$SCR(H) = N^{-1} \sum_{i=1}^N \{X_i - \hat{m}_H(\Theta_i)\}^2$$

Existe el problema que en el SCR el X_i es usado en el cálculo de $\hat{m}_H(\Theta_i)$ para predecirse a el mismo. Como consecuencia, la SCR puede hacerse tan pequeño como se quiera haciendo $H \rightarrow 0$. Esto conduce a la función de validación cruzada.

$$CV(H) = N^{-1} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2$$

En esta función se reemplaza $\hat{m}_H(\Theta_i)$ con el estimador

$$\hat{m}_{H,-i}(\Theta_i) = \frac{\sum_{j \neq i} K_H(\Theta_i - \Theta_j) X_j}{\sum_{j \neq i} K_H(\Theta_i - \Theta_j)}$$

lo cual es equivalente a otros enfoques que multiplican cada término por una función de penalización. Para el estimador de Nadaraya-Watson queda

$$CV(H) = \frac{1}{N} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2 = \frac{1}{N} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2 \left\{ \frac{X_i - \hat{m}_{H,-i}(\Theta_i)}{X_i - \hat{m}_H(\Theta_i)} \right\}^2$$

y

$$\begin{aligned} \frac{X_i - \hat{m}_H(\Theta_i)}{X_i - \hat{m}_{H,-i}(\Theta_i)} &= \frac{\sum_j K_H(\Theta_i - \Theta_j) X_j - X_i \sum_j K_H(\Theta_i - \Theta_j)}{\sum_{j \neq i} K_H(\Theta_i - \Theta_j) X_j - X_i \sum_{j \neq i} K_H(\Theta_i - \Theta_j)} \cdot \frac{\sum_{j \neq i} K_H(\Theta_i - \Theta_j)}{\sum_j K_H(\Theta_i - \Theta_j)} \\ &= 1 - \frac{K_H(0)}{\sum_j K_H(\Theta_i - \Theta_j)} \end{aligned}$$

Entonces el enfoque de validación cruzada es equivalente al de la función de penalización y comparte las mismas propiedades asintóticas. La validación cruzada es equivalente a la validación cruzada generalizada (Craven y Wahba, 1979). Härdle, Hall y Marron (1988) muestran la optimalidad asintótica del ancho de banda obtenido, aunque la tasa de convergencia es lenta. Una mejora en la selección del ancho de banda es discutida en Härdle, Hall y Marron (1992).

Métodos de selección de anchos de banda distintos a los de validación cruzada pueden ser encontrados en Fan y Gijbels (1996).

Capítulo 4

Modelo isótono

En este capítulo presentaremos un método no paramétrico para estimar las curvas características del ítem (CCI) cuando deben ser monótonas. Primero, se mostrará el caso unidimensional para luego generalizarlo al caso multidimensional.

Se trata de un procedimiento de dos etapas, en la primera etapa se utiliza un estimador no paramétrico de la curva característica del ítem mediante regresión no paramétrica basada en núcleos, y en la segunda etapa, usando el resultado anterior, se estimará la función de densidad de la inversa de la CCI.

Al integrar esta función de densidad obtenemos un estimador isótono de la inversa de la CCI, que se simetriza respecto a la bisectriz del cuadrado unidad para obtener el estimador de la CCI. También presentamos el caso multidimensional donde se procede coordenada a coordenada para obtener un estimador de la CCI isótono en cada dimensión.

En lo que se refiere a estimar funciones de regresión monótonas encontramos gran variedad de métodos como los de Cheng y Lin (1981), Wright (1981), Friedman y Tibshirani (1984), Delecroix y Thomas-Agnan (2000), Gijbels (2005) y Mammen (1991). Brunk (1955) propone un método que es modificado por Mukerjee (1988) para obtener estimadores con propiedades similares a los de regresión no paramétrica; Ramsay (1998), Kelly y Rice (1990) proponen utilizar splines. El procedimiento de Ramsay (1998) es semiparamétrico y es válido cuando la función de regresión satisface una ecuación diferencial de segundo orden por lo que es en general no consistente. Finalmente, Hall y Huang (2001) utilizan el estimador basado en núcleos con los pesos modificados y Dette, Neumeyer y Pilz (2006) utilizan un estimador en dos etapas.

Lee (2007) estudió el desempeño de tres métodos no paramétricos para estimar la CCI y compara el modelo de Ramsay con la regresión isótona y la regresión isótona suavizada. Los métodos de regresión isótona y de regresión isótona suavizada estiman la CCI bajo restricciones de monotonidad.

Bajo la restricción de monotonía de las CCI, Lee (2002) propuso el uso de estimaciones basadas en regresiones isótonas, las cuales son motivadas a partir de Barlow et al. (1972) y Robertson, Wright y Dykstra (1988). La regresión isótona es un método de mínimos cuadrados para ajustar datos bajo restricciones de orden. Afirmar que la función estimada es isótona implica que las funciones elegidas para ajustar los datos son funciones no decrecientes de la variable independiente. La regresión isótona para la estimación de la CCI la definen como sigue:

Definición 4.0.1. Consideremos en la habilidad de los examinados el siguiente orden simple $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$. Sea $P(\theta)$ una función dada de θ y sea F la familia de todas las funciones isótonas de θ . Entonces, $P^*(\theta)$ es la CCI estimada mediante regresión isótona si y solo si $P^*(\theta)$ es isótona y minimiza $\sum \left[P(\theta) - \hat{P}(\theta) \right]^2$, donde $\hat{P}(\theta)$ es un miembro de F . De aquí, $P^*(\theta)$ cumple

la restricción $P^*(\theta_1) \leq P^*(\theta_2) \leq \dots \leq P^*(\theta_k)$.

La solución más ampliamente usada para computar la regresión isótona es el algoritmo Pool-Adjacent-Violators (PAV) (Barlow et al., 1972; Hanson, Pledger and Wright, 1973). La solución de la regresión isótona a través del algoritmo de PAV se obtiene como sigue:

Paso 1: Ordenar los pares $(\theta_j, P(\theta_j))$ de todos los sujetos mediante la primera componente θ .

Paso 2: Se comienza con $P(\theta_{(1)})$, y se sigue en orden creciente o sea nos movemos a la derecha. Se para si el par $(P(\theta_{(j)}), P(\theta_{(j+1)}))$ viola la restricción de monotonía. Se toma $P(\theta_{(j)})$ y su adyacente $P(\theta_{(j+1)})$ y se remplazan ambos por el promedio:

$$P^*(\theta_{(j)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j)}) + P(\theta_{(j+1)})}{2}$$

Paso 3: Se verifica que $P(\theta_{(j-1)}) \leq P^*(\theta_{(j)})$. Si no es así, el conjunto $\{P(\theta_{(j-1)}), P(\theta_{(j)}), P(\theta_{(j+1)})\}$ se reemplaza por el promedio

$$P^*(\theta_{(j)}) = P^*(\theta_{(j-1)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j-1)}) + P(\theta_{(j)}) + P(\theta_{(j+1)})}{3}$$

Se continúa a la izquierda tanto como sea necesario, hasta que los requisitos de monotonía son satisfechos y entonces se continúa nuevamente a la derecha. La solución final es la estimación de la CCI isótona. Sin embargo, el método puede provocar que la función isótona resultante sea una función escalera si se analiza un pequeño grupo de examinados, o la monotonía en algún $P(\theta)$ es violada porque la probabilidad de las CCI fue remplazada por sus promedios. Bajo tales situaciones, la regresión isótona de las CCI son usualmente no muy suaves, pero no decrecientes.

Los estimados de las CCI mediante regresión isótona suavizada son obtenidos primero isotoniando los datos a través del algoritmo PAV y luego suavizando las funciones de regresión isótona usando un núcleo con un ancho de banda apropiado (Lee, 2002). La CCI resultante es generalmente más suave que la obtenida usando la estimación de la CCI mediante regresión isótona.

Si observamos detenidamente el procedimiento vemos que es esencialmente constructivo y nada garantiza la consistencia del estimador.

En este capítulo nos basaremos en el estimador propuesto por Dette, Neumeyer y Pilz (2006) para estimar regresiones monótonas en el intervalo $[0, 1]$. Consideramos que el método propuesto es una alternativa atractiva para estimar las CCI en casos en que estas deban ser crecientes, como por ejemplo en los contextos educativos.

Es un método de mucha flexibilidad, muy elegante y no iterativo. Esta es una ventaja importante respecto a los métodos paramétricos, tanto cuando se trata de modelos unidimensionales como en los modelos multidimensionales. Es justamente en el caso multidimensional donde aparece toda su potencia respecto a los métodos paramétricos ya que la cantidad de iteraciones en estos últimos es elevada. También se obtienen funciones suaves a partir del método original cosa que no es así en otros procedimientos no paramétricos.

Podemos mencionar como limitante que al trabajar en el intervalo $[0, 1]$ no obtenemos la distribución original del rasgo y la estimación de las CCI en el dominio de la habilidad. Sin embargo, esta dificultad es aparente dado que las escalas son equivalentes ante transformaciones monótonas. Si conocemos la distribución del rasgo en el caso unidimensional o las marginales en el caso multidimensional se obtendrán las escalas originales. Otra característica importante es que la dimensión del rasgo latente debe ser reducida ya que la cantidad de observaciones necesarias para una correcta estimación crece potencialmente con el número de dimensiones.

4.1. Modelo isótono

Como dijimos, nuestro enfoque se basa en el método desarrollado por Dette, Neumeyer y Pilz (2006). Mostraremos un estimador no paramétrico para las CCI cuando son isótonas. Como se explicó previamente, este es un supuesto habitual en TRI.

Dette et al. (2006) propusieron un método para estimar funciones monótonas a partir de estimadores no monótonos. Estos autores consideran funciones estrictamente monótonas en $[0, 1]$ con derivadas positivas.

Supongamos que la habilidad τ tiene distribución $F(\tau)$, y sea $P^*(\tau)$ la CCI la cual es monótona creciente y derivable. Notemos por $\theta = F(\tau)$, y $P(\theta) = P^*(F^{-1}(\theta)) = P^*(\tau)$. Es claro que $P(\theta)$ es creciente en $[0, 1]$, por lo tanto podemos suponer sin pérdida de generalidad que el rasgo tiene distribución uniforme en el intervalo unidad.

En este contexto, sean U_1, \dots, U_T una muestra de variables aleatorias con distribución uniforme en $[0, 1]$, K_d es un núcleo, y h_d una ventana. Entonces,

$$\frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(U_i) - u}{h_d}\right)$$

es el estimador de la densidad de $P(U)$.

Para $u \in [P(0), P(1)]$ tenemos que

$$P(P(U) \leq u) = P(U \leq P^{-1}(u)) = \int_0^{P^{-1}(u)} dt = \int_{P(0)}^u (P^{-1})'(z) dz$$

luego la densidad de $P(U)$ es $(P^{-1})'(u) \chi_{[P(0), P(1)]}(u)$. Entonces integrando,

$$\frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{P(U_i) - u}{h_d}\right) du$$

es un estimador consistente de P^{-1} en θ .

Es natural reemplazar $P(U_i)$ mediante un estimador de la CCI valorado en un conjunto de puntos igualmente espaciados en $[0, 1]$, lo cual realizaremos en nuestro caso mediante el modelo no paramétrico basado en el estimador de Nadaraya-Watson de regresión no paramétrica o en el modelo basado en el estimador mediante núcleos de polinomios locales de grado 1. Veamos el proceso con el estimador de Ramsay, entonces, en el contexto de nuestro enfoque, tomemos una grilla $0, \frac{1}{T}, \dots, \frac{i}{T}, \dots, 1$, y usemos el estimador no paramétrico de la regresión en cada punto

$$(4.1) \quad \widehat{P}\left(\frac{i}{T}\right) = \frac{\sum_{j=1}^N K_r\left(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r}\right)}$$

donde K_r y h_r son el núcleo y la ventana del estimador de regresión.

Entonces, el inverso de la CCI monótona en θ será

$$(4.2) \quad \widehat{P^{-1}}(\theta) = \frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}) - u}{h_d}\right) du$$

Finalmente, el estimador de \widehat{P} se obtiene por reflexión de $\widehat{P^{-1}}$ con respecto a $y = x$.

Para ilustrar el caso multidimensional veremos como proceder en dos dimensiones. Supongamos una CCI creciente en cada rasgo y que el estimador de la CCI se obtuvo como se vio previamente usando regresión no paramétrica en dos dimensiones.

$$(4.3) \quad \widehat{P}_i(\theta_1, \theta_2) = \frac{\sum_{k=1}^N K_r\left(\frac{\widehat{\theta}_{1i} - \theta_1}{h_1}, \frac{\widehat{\theta}_{2i} - \theta_2}{h_2}\right) X_{ik}}{\sum_{k=1}^N K_r\left(\frac{\widehat{\theta}_{1k} - \theta_1}{h_1}, \frac{\widehat{\theta}_{2k} - \theta_2}{h_2}\right)}$$

donde $e K_r$ es un núcleo de dimensión dos, con soporte compacto $C \subset [0, 1]^2$. También suponemos que tanto el núcleo, las ventanas y las funciones utilizadas para estimar los rasgos cumplen las hipótesis requeridas para que este estimador sea consistente.

Tomemos una grilla $0, \frac{1}{T}, \dots, \frac{j}{T}, \dots, 1$ para θ_1 y una grilla $0, \frac{1}{T}, \dots, \frac{j}{T}, \dots, 1$ para θ_2

Para cada $\theta_2 \in (0, 1)$ fijo, consideremos el estimador

$$(4.4) \quad \widehat{H}^{-1}(\theta_1|\theta_2) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}, \theta_2) - u}{h_d}\right) du$$

La función $\widehat{H}^{-1}(\theta_1|\theta_2)$ es estrictamente creciente en θ_1 para cada θ_2 fijo.

Podemos calcular el inverso (como función de θ_1) para obtener $\widehat{H}(\theta_1|\theta_2)$.

Para cada $\theta_1 \in (0, 1)$ fijo, calculamos

$$(4.5) \quad \widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^T \int_{-\infty}^{\theta_2} K_d\left(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}\right) du$$

Finalmente, $\widehat{P^*}(\theta_1, \theta_2)$ se calcula mediante la inversa (como función de θ_2).

El algoritmo sería:

Paso 1 Usando el procedimiento estandar para cada $1 \leq i \leq T$, y $1 \leq j \leq T$, calcular

$$\widehat{P}\left(\frac{i}{T}, \frac{j}{T}\right)$$

Paso 2 Para cada $\frac{j}{T}$ con $1 \leq j \leq T$, calcular $H^{-1}(\theta_1|\frac{j}{T})$ mediante

$$(4.6) \quad \widehat{H}^{-1}(\theta_1|\frac{j}{T}) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}, \frac{j}{T}) - u}{h_d}\right) du$$

Paso 3 Invirtiendo con respecto a θ_1 , obtenemos $\widehat{H}(\theta_1|\frac{j}{T})$ with $1 \leq j \leq T$

Paso 4 Calcular

$$\widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^T \int_{-\infty}^{\theta_2} K_d\left(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}\right) du$$

Paso 5 Invirtiendo respecto a θ_2 , obtenemos $\widehat{P^*}(\theta_1, \theta_2)$.

4.2. Propiedades del estimador cuando el rasgo es medido sin error

Veremos las propiedades del estimador en el caso más simple donde se puede aplicar directamente el modelo de regresión. Supondremos que el rasgo es conocido o medido sin error y sin pérdida de generalidad que las respuestas son continuas en $[0, 1]$ y se quiere estimar la esperanza condicional de las respuestas con respecto al rasgo. Trabajaremos en una dimensión pues la extensión de los resultados a una dimensión mayor es clara. Para mantener la notación pondremos $E(X_i|\Theta) = P(\Theta)$. Consideremos el modelo

$$X_i = P(\Theta_i) + \sigma(\Theta_i)\varepsilon_i, \quad i = 1, \dots, N$$

donde $\{\Theta_i, X_i\}$ $i = 1, \dots, N$ es una muestra bivariada de observaciones independientes idénticamente distribuidas tales que Θ_i tiene densidad f dos veces continuamente diferenciable, con soporte compacto $[0, 1]$. Asumiremos que las variables aleatorias ε_i son i.i.d. con $E(\varepsilon_i) = 0$,

$E(\varepsilon_i^2) = 1$ y momento cuarto finito. También asumiremos que la función $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ y la función de P son continuas y dos veces continuamente diferenciables, respectivamente.

Vimos más arriba que si $\Theta_i \sim f$ entonces $\frac{1}{Th_d} \sum_{i=1}^T K_d(\frac{\Theta_i - \theta}{h_d}) = \hat{f}(\theta)$ es un estimador de la densidad f y

$$(4.7) \quad \widehat{P_I^{-1}}(\theta) = \int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T}) - u}{h_d}\right) du$$

Definamos

$$(4.8) \quad P_T^{-1}(\theta) = \int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T}) - u}{h_d}\right) du$$

Supondremos que K_d y K_r tienen soporte compacto, por ejemplo, sin pérdida de generalidad en $[-1, 1]$.

Hay dos resultados principales uno sobre una aproximación de $P_T(\theta)$ a $P(\theta)$ y otro que obtiene la distribución asintótica.

El primer resultado establece que si las hipótesis mencionadas se cumplen y además P es estrictamente creciente entonces para todo $\theta \in [0, 1]$ con $P'(\theta) > 0$

$$(4.9) \quad P_T(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Th_d}\right)$$

donde $\kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv$

El resultado asintótico establece que si \hat{P}_I y P_T denotan las inversas de las funciones \hat{P}_I^{-1} y P_T^{-1} . Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$, existe entonces tenemos, que para $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{Nh_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{f} \right) (\theta) \right) \Rightarrow N(0, s^2(\theta))$$

donde la varianza asintótica está dada por

$$s^2(\theta) = \frac{\sigma^2(\theta)P'(\theta)}{f(\theta)} \int \int \int K_d(w + cP'(\theta)(v - u)) K_d(w) K_r(u) K_r(v) dw du dv$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$ tenemos que para todo $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{Nh_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{f} \right) (\theta) \right) \Rightarrow N(0, \tilde{s}^2(\theta))$$

donde la varianza asintótica está dada por $\tilde{s}^2(\theta) = \frac{\sigma^2(\theta)}{f(\theta)} \int K_r^2(u) du$.

Observación 4.2.1. $\widehat{P_I^{-1}} = 1$ si $\theta > \max_{i=1}^T \hat{P}(\frac{i}{T}) + h_d$ y $\widehat{P_I^{-1}} = 0$ si $\theta < \min_{i=1}^T \hat{P}(\frac{i}{T}) - h_d$

Dado que $\widehat{P_I^{-1}} = \sum_{i=1}^T \int_{-\infty}^{\theta} \frac{1}{Th_d} K_d\left(\frac{\hat{P}(\frac{i}{T}) - u}{h_d}\right) du$ como K_d tiene soporte $[-1, 1]$ entonces $K_d = 0$

si para todo i se cumple alguna de las desigualdades $\begin{cases} \frac{\hat{P}(\frac{i}{T}) - u}{h_d} > 1 \\ \frac{\hat{P}(\frac{i}{T}) - u}{h_d} < -1 \end{cases}$ o lo que es lo mismo $\hat{P}(\frac{i}{T}) - h_d > u$, o $\hat{P}(\frac{i}{T}) + h_d < u$ para todo i .

Si $\theta < \min_{i=1}^T \hat{P}(\frac{i}{T}) - h_d$, para todo $u \in (-\infty, \theta]$ y para todo i se cumple $u < \hat{P}(\frac{i}{T}) - h_d$ entonces $\hat{P}_I^{-1}(\theta) = 0$. Si $\theta > \max_{i=1}^T \hat{P}(\frac{i}{T}) + h_d$ entonces tenemos que

$$\int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du = \int_{-\infty}^{\min \hat{P}(\frac{i}{T})-h_d} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du +$$

$$\int_{\min \hat{P}(\frac{i}{T})-h_d}^{\max \hat{P}(\frac{i}{T})+h_d} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du + \int_{\max \hat{P}(\frac{i}{T})+h_d}^{+\infty} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du = 1$$

ya que la primera y última integral valen 0 por verificarse las desigualdades vistas más arriba y la segunda integral vale 1, ya que el núcleo tiene soporte $[-1, 1]$.

Observación 4.2.2. Tenemos que $P_T^{-1}(\theta) = \int_0^1 \chi_{\{P(x) \leq \theta\}} dx + o(1)$ dado que

$$\int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) du \rightarrow \int_{-\infty}^{\theta} (P^{-1}(u))' du = P^{-1}(\theta) = \int_0^{P^{-1}(\theta)} d\theta =$$

$$\int_0^1 \chi_{\{P^{-1}(\theta) \geq x\}} dx = \int_0^1 \chi_{\{P(x) \leq \theta\}} dx$$

Observación 4.2.3. También se cumple que $\int_0^1 \chi_{\{P(x) \geq \theta\}} dx = P^{-1}(\theta) = \inf\{u : P(u) > \theta\}$

Observación 4.2.4. P_T^{-1} es estrictamente creciente si $\max_{i=1..T-1} [P(\frac{i+1}{T}) - P(\frac{i}{T})] < 2h_d$.

Sea $\theta_2 > \theta_1$ $P_T(\theta_2) - P_T(\theta_1) = \frac{1}{Th_d} \int_{\theta_1}^{\theta_2} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) du$ y este término es positivo excepto si $K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) = 0$ para todo i y $u \in (\theta_1, \theta_2)$. Como K_d tiene soporte $[-1, 1]$ entonces esto se verifica si $|\frac{P(\frac{i}{T})-u}{h_d}| > 1 \forall i$ y $u \in (\theta_1, \theta_2)$ o sea $|P(\frac{i}{T}) - u| > h_d$.

Si $\max_{i=1..T-1} [P(\frac{i+1}{T}) - P(\frac{i}{T})] < 2h_d$, entonces, si $u \in [P(\frac{i}{T}), \frac{P(\frac{i+1}{T})+P(\frac{i}{T})}{2}]$, entonces, $|P(\frac{i}{T}) - u| < h_d$ y si $u \in [\frac{P(\frac{i}{T})+P(\frac{i+1}{T})}{2}, P(\frac{i+1}{T})]$, entonces $|P(\frac{i+1}{T}) - u| < h_d$ luego; $\forall u \in (\theta_1, \theta_2)$, existe i tal que $K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) > 0$.

Igualmente \hat{P}_I^{-1} es estrictamente creciente si $\max_{i=1..T-1} [\hat{P}(\frac{i+1}{T}) - \hat{P}(\frac{i}{T})] < 2h_d$.

Las condiciones anteriores se cumplen para T y N suficientemente grandes por la continuidad de P y su estimación \hat{P} .

Consideremos para hacer más simple la notación que el número T de puntos en la grilla es igual al tamaño muestral N por lo que de ahora en adelante $T=N$

Lema 4.2.5. Si el núcleo K es acotado y P es derivable entonces

$$(4.10) \quad P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx + O\left(\frac{1}{Th_d}\right)$$

Demostración. Como $T=N$ sea $g(N, h) = \int_{-\infty}^{\theta} \frac{1}{h_d} \sum_{i=1}^N \frac{1}{N} K\left(\frac{P(\frac{i}{N})-u}{h_d}\right) du - \int_{-\infty}^{\theta} \frac{1}{h_d} \int_0^1 K\left(\frac{P(x)-u}{h_d}\right) dx du$ luego tenemos

$$(4.11) \quad Ng(N, h) = \int_{-\infty}^{\theta} \frac{1}{h_d} \sum_{i=1}^N K\left(\frac{P(\frac{i}{N})-u}{h_d}\right) du - N \int_{-\infty}^{\theta} \frac{1}{h_d} \int_0^1 K\left(\frac{P(x)-u}{h_d}\right) dx du$$

Como $\int_{-\infty}^{\theta} \frac{1}{h_d} K\left(\frac{P(x)-u}{h_d}\right) du = \int_{P(x)-h_d}^{\theta} \frac{1}{h_d} K\left(\frac{P(x)-u}{h_d}\right) du = \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(z) dz$ entonces 4.11 se transforma en

$$(4.12) \quad \sum_{i=1}^N \left[\int_0^1 \int_{-1}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z) dz - \int_0^1 \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(z) dz dx \right] \leq \sum_{i=1}^N \int_0^1 \int_{\frac{\theta-P(x)}{h_d}}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z) dz$$

Como $K(z)$ está acotado por C y usando 4.12 resulta que:

$$\begin{aligned} \int_{\frac{\theta-P(x)}{h_d}}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z)dz &\leq \frac{C}{h_d}(P(\frac{i}{N})-P(x)) \leq \frac{C}{h_d} \sum_{i=1}^N \int_0^1 (P(\frac{i}{N})-P(x))dx = \frac{C}{h_d} \sum_{i=1}^N \int_0^1 C_1((\frac{i}{N})-x)dx \leq \\ &\leq \frac{CC_1}{h_d} \sum_{i=1}^N ((\frac{i}{N}) - \frac{1}{2}) = \frac{C_2}{h_d} (\frac{N(N+1)}{2N} - N\frac{1}{2}) = \frac{C_2}{2h_d} \end{aligned}$$

Entonces $Nhg(N, h) \leq C_2$ o sea $g(N, h) = O(\frac{1}{Nh})$

Si tomamos N y T cualesquiera también

$$P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx + O(\frac{1}{Th_d})$$

Lema 4.2.6. Si P es estrictamente creciente y las hipótesis ya vistas entonces:

para todo $\theta \in (P(0), P(1))$ con $(P^{-1})'(\theta) > 0$

$$(4.13) \quad P_T^{-1}(\theta) = P^{-1}(\theta) + \kappa_2(K_d)h_d^2(P^{-1})'' + o(h_d^2) + O(\frac{1}{Th_d})$$

$$\text{con } \kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv.$$

Demostración. K_d tiene soporte $[-1, 1]$ usando el lema anterior debe ser entonces

$-1 \leq \frac{P(x)-u}{h_d} \leq 1$ o sea $u \geq P(x)-h_d$ y $u \leq P(x)+h_d$. Para ese x se cumple que $P(x)-h_d = \theta$ lo que implica que varia entre 0 y $P^{-1}(\theta+h_d)$ y u varia entre $P(x)-h_d$ y θ , entonces llamando

$$\begin{aligned} A(h_d) &= \int_0^1 \int_{-\infty}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx = \int_0^{P^{-1}(\theta+h_d)} \int_{P(x)+h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx = \\ &= \int_0^{P^{-1}(\theta-h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx + \int_{P^{-1}(\theta-h_d)}^{P^{-1}(\theta+h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx \end{aligned}$$

En la primera integral haciendo el cambio de variable $-w = \frac{P(x)-u}{h_d}$ y luego aplicando partes

$$\begin{aligned} \int_0^{P^{-1}(\theta-h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx &= \int_0^{P^{-1}(\theta-h_d)} \int_{-1}^{\frac{-P(x)+\theta}{h_d}} K_d(w) dw dx = \\ (4.14) \quad &= x \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(w) dw \Big|_0^{P^{-1}(\theta-h_d)} + \int_0^{P^{-1}(\theta+h_d)} x K_d(\frac{\theta-P(x)}{h_d}) \frac{P'(x)}{h_d} dx \end{aligned}$$

donde la última integral vale cero ya que $x \in [0, P^{-1}(\theta-h_d)]$.

Recordemos que $K_d \neq 0$ si $|\frac{\theta-P(x)}{h_d}| < 1$ o sea $P(x) \in [\theta-h_d, \theta+h_d]$. Como P es monótona debe ser $x \in [P^{-1}(\theta-h_d), P^{-1}(\theta+h_d)]$ pero en nuestro caso $x \in [0, P^{-1}(\theta-h_d)]$, entonces $K_d = 0$ luego la última integral vale cero.

La segunda integral de 4.14 es igual a:

$$\begin{aligned} \int_{P^{-1}(\theta-h_d)}^{P^{-1}(\theta+h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx &= \\ \int_0^1 \chi\{P^{-1}(\theta-h_d) \leq x \leq P^{-1}(\theta+h_d)\} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx \end{aligned}$$

Haciendo el cambio de variables $v = \frac{P(x)-u}{h_d}$ resulta que:

$$\begin{aligned} \int_0^1 \chi\{P^{-1}(\theta - h_d) \leq x \leq P^{-1}(\theta + h_d)\} \int_1^{\frac{P(x)-\theta}{h_d}} -K_d(v)dvdx = \\ = \int_0^1 \chi\{P^{-1}(\theta - h_d) \leq x \leq P^{-1}(\theta + h_d)\} \int_{\frac{P(x)-\theta}{h_d}}^1 K_d(v)dvdx \end{aligned}$$

Haciendo el cambio de variables $z = \frac{P(x)-\theta}{h_d}$ en las indicatrices tenemos:

$$\int_{\frac{P(0)-\theta}{h_d}}^{\frac{P(1)-\theta}{h_d}} \chi\{-1 \leq z \leq 1\} \int_z^1 h_d K_d(v) (P^{-1})'(\theta + h_d z) dz = h_d \int_{-1}^1 \int_z^1 K_d(v) (P^{-1})'(\theta + h_d z) dz$$

Entonces, $A(h_d) = P^{-1}(\theta - h_d) + h_d \int_{-1}^1 (P^{-1})'(\theta + h_d z) \int_z^1 K_d(v) dv dz$. Aplicamos Taylor y

$$P^{-1}(\theta - h_d) = P^{-1}(\theta) - (P^{-1})'(\theta)(h_d) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2)$$

$$(P^{-1})'(\theta + zh_d) = (P^{-1})'(\theta) + (P^{-1})''(\theta)(zh_d) + o(h_d)$$

Remplazando resulta $A(h_d) = P^{-1}(\theta) - (P^{-1})'(\theta)(h_d) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2) +$

$$h_d \int_{-1}^1 [(P^{-1})'(\theta) + (P^{-1})''(\theta)(zh_d) + o(h_d)] \int_z^1 K_d(v) dv dz$$

Como $\int_{-1}^1 \int_z^1 K_d(v) dv dz = 1$ (observación 4.2.7) resulta $h_d \int_{-1}^1 (P^{-1})'(\theta) \int_z^1 K_d(v) dv dz = h_d (P^{-1})'(\theta)$. Entonces,

$$A(h_d) = P^{-1}(\theta) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2) + h_d \int_{-1}^1 (P^{-1})''(\theta) zh_d \int_z^1 K_d(v) dv dz + h_d(o(h_d)) \int_{-1}^1 \int_z^1 K_d(v) dv dz$$

$$\text{luego } A(h_d) = P^{-1}(\theta) + h_d^2 (P^{-1})''(\theta) \left(\frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz \right) + o(h_d^2)$$

Apliquemos partes a:

$$\int_{-1}^1 z \int_z^1 K_d(v) dz = \frac{z^2}{2} \int_z^1 K_d(v) dv \Big|_{-1}^1 + \int_1^1 \frac{z^2}{2} K_d(z) dz = \frac{-1}{2} + \int_{-1}^1 \frac{z^2}{2} K_d(z) dz =$$

$$\text{entonces obtenemos } \int_{-1}^1 \frac{z^2}{2} K_d(z) dz = \frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz.$$

$$\text{Ahora llamamos } \kappa_2(K_d) = \int_{-1}^1 \frac{z^2}{2} K_d(z) dz = \frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz \text{ luego}$$

$$A(h_d) = P^{-1}(\theta) + P^{-1}''(\theta) h_d^2 \kappa_2(K_d) + o(h_d^2)$$

Por el lema 4.2.5 sabemos que

$$P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K_d\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx + O\left(\frac{1}{Th_d}\right) = P^{-1}(\theta) + P^{-1}''(\theta) h_d^2 \kappa_2(K_d) + o(h_d^2) + O\left(\frac{1}{Th_d}\right)$$

lo que prueba el lema.

Observación 4.2.7. Probemos aplicando partes que $\int_{-1}^1 \int_z^1 K_d(v) dv dz = 1$

$$\int_{-1}^1 K_d(v) dv dz = 1 = z \int_z^1 K_d(v) dv \Big|_{-1}^1 - \int_{-1}^1 z(-K_d(z)) dz = 1 - K_d(1) \int_1^1 z dz + \int_{-1}^1 z K_d(z) dz = 1$$

Definición 4.2.8. Sea θ fijo. Definamos $M = \{H \in C^2[0, 1] : H' > 0 \text{ en } [0, 1] \text{ el cual contiene a } \theta \text{ en el interior de la imagen, es decir } \theta \in \text{int}(H[0, 1])\}$

Definición 4.2.9. Definimos $\phi : M \rightarrow [0, 1]$ dada por $\phi(H) = H^{-1}(t)$.

También definimos para $H_1, H_2 \in M$ $Q[0, 1] \rightarrow \mathbb{R}$ dada por $\lambda \rightarrow \phi(H_1 + \lambda(H_2 - H_1))$

o sea $Q(\lambda) = \phi(H_1 + \lambda(H_2 - H_1)) = (H_1 + \lambda(H_2 - H_1))^{-1}(t)$

Lema 4.2.10. El mapa $Q : [0, 1] \rightarrow \mathbb{R}$ es dos veces diferenciable con

$$Q'(\lambda) = \frac{H_2 - H_1}{H'_1 + \lambda(H'_2 - H'_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t)$$

$$Q''(\lambda) = Q'(\lambda) \cdot \left\{ -2 \frac{H'_2 - H'_1}{H'_1 + \lambda(H'_2 - H'_1)} + \frac{(H_2 - H_1)(H''_1 + \lambda(H''_2 - H''_1))}{(H'_1 + \lambda(H'_2 - H'_1))^2} \right\} \circ Q(\lambda)$$

Demostración. Sea $F(x, y) = (H_1 + x(H_2 - H_1))(y) - t$ entonces

$$F(\lambda, Q(\lambda)) = (H_1 + \lambda(H_2 - H_1))(H_1 + \lambda(H_2 - H_1))^{-1}(t) - t = t - t = 0$$

Es fácil ver que el dominio de Q se puede extender a un entorno de $[0, 1]$ entonces aplicamos el teorema de la Función Implícita.

Definimos $y = Q(x)$, o sea, consideramos $F(x, Q(x))$. En λ tenemos que $F(\lambda, Q(\lambda)) = 0$ entonces existe $Q'(\lambda)$ y se puede calcular como $\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{-F_x}{F_y}$ evaluado en $(\lambda, Q(\lambda))$.

Resulta $\frac{\partial F}{\partial x} = (H_2 - H_1)(y)$ y también $\frac{\partial F}{\partial y} = \frac{\partial H_1}{\partial y} + x \frac{\partial (H_2 - H_1)(y)}{\partial y} = H'_1(y) + x(H'_2 - H'_1)(y)$

Entonces $\frac{\partial Q(x)}{\partial x}(\lambda) = -\frac{(H_2 - H_1)(y)}{H'_1(y) + x(H'_2 - H'_1)(y)} \mid_{(\lambda, Q(\lambda))}$

$$Q'(\lambda) = -\frac{(H_2 - H_1)(Q(\lambda))}{H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))} = -\frac{(H_2 - H_1)}{H'_1 + \lambda(H'_2 - H'_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t)$$

Para la derivada segunda derivamos la derivada primera. A partir de

$$Q'(\lambda) = -\frac{(H_2 - H_1)(Q(\lambda))}{H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))}$$

La derivada del numerador es: $[(H_2 - H_1)(Q(\lambda))]' = (H'_2 - H'_1)(Q(\lambda))Q'(\lambda)$, la derivada del denominador es:

$$[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]' = H''_1(Q(\lambda))Q'(\lambda) + (H'_2 - H'_1)(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))Q'(\lambda)$$

$$\text{Entonces, } Q''(\lambda) = -\left[\frac{(H'_2 - H'_1)(Q(\lambda))Q'(\lambda)(H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda)))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} - \frac{(H_2 - H_1)(Q(\lambda))(H''_1(Q(\lambda))Q'(\lambda) + (H'_2 - H'_1)(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))Q'(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} \right]$$

$$\text{Agrupando resulta que: } Q''(\lambda) = \frac{(H_2 - H_1)(Q(\lambda))(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} - \frac{Q'(\lambda)[((H'_2 - H'_1)(H'_1 + \lambda(H'_2 - H'_1))(Q(\lambda)) - (H_2 - H_1)(Q(\lambda))[H''_1(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))])]}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2}$$

Entonces tenemos que:

$$Q''(\lambda) = \frac{-Q'(\lambda)(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]} - \frac{Q'(\lambda)(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]} + \frac{Q'(\lambda)(H_2 - H_1)(Q(\lambda))[H''_1(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))]}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2}$$

$$\text{Finalmente resulta } Q(\lambda)'' = Q'(\lambda) \left(\frac{-2(H'_2 - H'_1)}{[H'_1 + \lambda(H'_2 - H'_1)]} + \frac{[H_2 - H_1][H''_1 + \lambda(H''_2 - H''_1)]}{[H'_1 + \lambda(H'_2 - H'_1)]^2} \right) (Q(\lambda))$$

Lema 4.2.11. Si P es estrictamente creciente y las hipótesis anteriores se cumplen entonces para todo $\theta \in [0, 1]$ con $P'(\theta) > 0$

$$(4.15) \quad P_T(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Th_d}\right)$$

donde $\kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv$

Demostración. Como estamos suponiendo que $T=N$ probaremos

$$P_N(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Nh_d}\right)$$

Por definición tenemos que $P_N(\theta) = \phi(P_N^{-1}(\theta))$ entonces,

$$P_N(\theta) - P(\theta) = \phi(P_N^{-1}(\theta)) - \phi(P^{-1}(\theta)) = Q(1) - Q(0) = Q'(\lambda^*) \text{ con } \lambda^* \in [0, 1].$$

Donde hemos considerado $Q(\lambda) = \phi(P^{-1} + \lambda(P_N^{-1} - P^{-1}))$ tomando $H_1 = P^{-1}$, $H_2 = P_N^{-1}$.

Como $Q'(\lambda) = -\frac{(H_2 - H_1)}{H_1' + \lambda(H_2' - H_1')} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(\theta)$ en nuestro caso tenemos que

$$Q'(\lambda^*) = -\frac{(P_N^{-1} - P^{-1})}{P^{-1'} + \lambda^*(P_N^{-1} - P^{-1})'} \circ (P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))^{-1}(\theta)$$

Por el lema 4.2.6 $P^{-1} + \lambda^*(P_N^{-1} - P^{-1}) \rightarrow P^{-1}$, entonces, $\theta_N = (P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))^{-1}(\theta) \rightarrow P(\theta)$

Aplicando el teorema del valor medio

$$(P_N^{-1} - P^{-1})(\theta_N) - (P_N^{-1} - P^{-1})(P(\theta)) = (P_N^{-1} - P^{-1})'(\eta)(\theta_N - P(\theta))$$

con $\eta - P(\theta) \leq \theta_N - P(\theta)$. Ahora

$$\begin{aligned} (P_N^{-1} - P^{-1})'(\eta) &= (P_N^{-1})'(\eta) - (P^{-1})'(\eta) = \left[\int_{-\infty}^{\eta} \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-u}{h_d}\right) dx du + O\left(\frac{1}{Nh_d}\right) \right]'(\eta) - P^{-1'}(\eta) = \\ &= \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-\eta}{h_d}\right) dx - P^{-1'}(\eta) + O\left(\frac{1}{Nh_d}\right) \end{aligned}$$

A partir de la igualdad

$$\int_0^1 \int_{-\infty}^{\eta} K_d\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx = P^{-1}(\eta) + P^{-1''}(\theta)h_d^2\kappa_2(K_d) + o(h_d^2)$$

obtenemos $\int_0^1 K_d\left(\frac{P(x)-\eta}{h_d}\right) \frac{1}{h_d} dx = (P^{-1})'(\eta) + O(h_d^2)$ reemplazando queda

$$Q'(\lambda^*) = -\frac{(P_N^{-1} - P^{-1})(P(\theta)) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{(P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))'(\theta_N)}. \text{ Por el lema anterior tenemos que,}$$

$$P_N^{-1} - P^{-1} = \kappa_2(K_d)h_d^2(P^{-1}(\theta))'' + o(h_d^2) + O\left(\frac{1}{Nh_d}\right) \text{ entonces}$$

$$Q'(\lambda) = -\frac{\kappa_2(K_d)h_d^2(P^{-1}(\theta))'' + o(h_d^2) + O(\frac{1}{Nh_d}) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{(P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))'(\theta_N)} \text{ de donde}$$

$$Q'(\lambda) = \frac{\kappa_2(K_d)h_d^2 \frac{P''(\theta)}{P'(\theta)^3} + o(h_d^2) + O(\frac{1}{Nh_d}) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{\frac{1}{P'(\theta)}}$$

donde se ha utilizado que $(P^{-1}(P(\theta)))'' = -\frac{P''(\theta)}{(P'(\theta))^3}$.

Como $\lim_N \frac{O(h_d^2)(\theta_N - P(\theta))}{h_d^2} = 0$ y $\lim_N \frac{O(\frac{1}{Nh_d})(\theta_N - P(\theta))}{\frac{1}{Nh_d}} = 0$, ya que $\theta_N \rightarrow P(\theta)$ obtenemos el resultado.

Asumiremos ahora que $h_r \rightarrow 0$, $h_d \rightarrow 0$, $Nh_r \rightarrow \infty$, $Nh_d \rightarrow \infty$ y adicionalmente que

$$Nh_r^5 = O(1), \quad n = O(T), \quad \frac{\log h_r^{-1}}{Nh_r h_d^3} = o(1)$$

Teorema 4.2.12. Si las condiciones anteriores son satisfechas, $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$ existe y P es estrictamente creciente; entonces, para todo $\theta \in (P(0), P(1))$ con $P'(P^{-1}(\theta)) > 0$,

$$(4.16) \quad \sqrt{Nh_d}(\widehat{P}_I^{-1}(\theta) - P_T^{-1}(\theta) + \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{Pf'} \right)(P^{-1}(\theta))) \Rightarrow N(0, r^2(\theta))$$

donde la constante $\kappa_2(K_r)$ esta definida en el lema 4.2.6 y la varianza asintótica está dada por

$$(4.17) \quad r^2(\theta) = \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \times \int \int \int K_d(w + cP'(P^{-1}(\theta))(v - u))K_d(w)K_r(v)dw dv du$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$, entonces tenemos, para todo $\theta \in (P(0), P(1))$ con $P'(P^{-1}(\theta)) > 0$,

$$\sqrt{N h_d}(\hat{P}_I^{-1}(\theta) - P_T^{-1}(\theta) + \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{P f'})(P^{-1}(\theta))) \Rightarrow N(0, \bar{r}^2(\theta))$$

donde la varianza asintótica está dada por

$$\bar{r}^2(\theta) = \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int K_r^2(u)du$$

Demostración. Usando la descomposición

$$(4.18) \quad \hat{P}_I^{-1}(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^N K_d(\frac{\hat{P}(\frac{i}{N})-u}{h_d})du = P_N^{-1}(\theta) + \Delta_N(\theta)$$

donde hemos sumado y restado $P_T^{-1}(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^T K_d(\frac{P(\frac{i}{N})-u}{h_d})du$ y tomado

$$(4.19) \quad \Delta_N(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^N \left\{ K_d(\frac{\hat{P}(\frac{i}{N})-u}{h_d}) - K_d(\frac{P(\frac{i}{N})-u}{h_d}) \right\} du$$

Ahora descomponemos Δ_N en

$$(4.20) \quad \Delta_N(\theta) = \Delta_N^1(\theta) + \frac{1}{2} \Delta_N^2(\theta)$$

donde,

$$\Delta_N^1(\theta) = \frac{1}{N h_d^2} \int_{\infty}^{\theta} \sum_{i=1}^N K_d'(\frac{P(\frac{i}{N})-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \} du$$

$$\Delta_N^2(\theta) = \frac{1}{N h_d^3} \int_{\infty}^{\theta} \sum_{i=1}^N K_d''(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 du$$

con $|\xi_i - P(\frac{i}{N})| < |\hat{P}(\frac{i}{N}) - P(\frac{i}{N})|$ ($i = 1, \dots, n$) donde hemos usado que $f(x) = f(c) + f'(c)(x - c) + f''(\xi)(x - c)^2$. Luego

$$\Delta_N^2(\theta) = \frac{1}{N h_d^3} \sum_{i=1}^N \int_{\infty}^{\theta} K_d''(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 du = \frac{1}{N h_d^3} \sum_{i=1}^N \int_{\infty}^{\theta} K_d''(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 \frac{du}{h_d} =$$

$$\frac{1}{N h_d^2} \left| \sum_{i=1}^N K_d'(\frac{\xi_i-u}{h_d}) \right|_{\infty}^{\theta} \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 = \frac{1}{h_d^2} \left| \frac{1}{N} \sum_{i=1}^N K_d'(\frac{\xi_i-\theta}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 \right| =$$

$$= \frac{1}{h_d^2} \left| \int_0^1 K_d'(\frac{P(x)-\theta}{h_d}) \{ \hat{P}(x) - P(x) \}^2 dx \right| \cdot (1 + o_p(1))$$

Si asumimos que el núcleo K_r ha sido modificado cerca de los bordes entonces tenemos que este término es de orden $O((\frac{1}{N h_r} + h_r^4)/h_d)$. Esto implica que

$$(4.21) \quad \sqrt{N h_d} \Delta_N^2(\theta) = o_p(1),$$

ya que,

$$\lim_{n \rightarrow \infty} \sqrt{N h_d} \left(\frac{1}{N h_r} + h_r^4 \right) (1 + o_p(1)) = \lim_{n \rightarrow \infty} \sqrt{N h_d} \left(\frac{1 + N h_r^5}{N h_r h_d} \right) O_p(1) = \lim_{n \rightarrow \infty} \frac{O_p(1)}{\sqrt{N h_d h_r}} =$$

$$\lim_{n \rightarrow \infty} \frac{O_p(1)}{n^{\frac{1}{2}} h_r^{\frac{3}{2}}} = \frac{O_p(1) h_r}{(N h_r^5)^{\frac{1}{2}}} = o_p(1).$$

Combinando 4.18, 4.20 y 4.21 muestra que la afirmación del teorema puede ser probada demostrando la convergencia débil

$$(4.22) \quad \sqrt{N h_d} \left(\Delta_N^1(\theta) + \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{P f'} \right) (P^{-1}(\theta)) \right) \xrightarrow{D} N(0, r^2(\theta))$$

Para esto usamos la descomposición

$$\Delta_N^1(\theta) = (\Delta_N^{(1,1)}(\theta) + \Delta_N^{(1,2)}(\theta))(1 + o_p(1))$$

con

$$\begin{aligned} \Delta_N^{(1,1)}(\theta) &= \frac{-1}{n^2 h_d h_r} \sum_{i,j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{P(\Theta_j) - P(\frac{i}{N})}{f(\frac{i}{N})} \\ \Delta_N^{(1,2)}(\theta) &= \frac{-1}{n^2 h_d h_r} \sum_{i,j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{\sigma(\Theta_j) \epsilon_j}{f(\frac{i}{N})} \end{aligned}$$

ya que, $X_i = P(\Theta_i) + \sigma(\Theta_i) \epsilon_i$, $\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \rightarrow f$ y $\hat{P}(\Theta) = \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \Theta}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \Theta}{h_r}\right)}$ luego

$$\begin{aligned} (\Delta_N^{(1)}(\theta)) &= \frac{1}{n h_d^2} \sum_{i=1}^N \int_0^1 K_d'\left(\frac{P(\frac{i}{N}) - u}{h_d}\right) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) du = \\ &= \frac{1}{n h_d} \sum_{i=1}^N (-K_d\left(\frac{P(\frac{i}{N}) - u}{h_d}\right)|_{\infty}^{\theta}) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) \\ &= \frac{-1}{n h_d} \sum_{i=1}^N (K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right)) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) \end{aligned}$$

Como

$$\hat{P}\left(\frac{i}{N}\right) = \frac{\sum_{j=1}^N K_r\left(\frac{\frac{i}{N} - \Theta_j}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\frac{i}{N} - \Theta_j}{h_r}\right)}$$

Entonces,

$$\hat{P}\left(\frac{i}{N}\right) - P\left(\frac{i}{N}\right) = \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (X_j - P(\frac{i}{N}))}{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right)} = \frac{1}{N h_r} \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{\frac{1}{N h_r} \sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right)}$$

Entonces tenemos que

$$\begin{aligned} \Delta_N^1(\theta) &= -\frac{1}{n^2 h_r h_d} \frac{\sum_{i=1}^N \sum_{j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{\hat{f}\left(\frac{i}{N}\right)} \\ \Delta_N^1(\theta) &= -\frac{1}{n^2 h_r h_d} \frac{\sum_{i=1}^N \sum_{j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{f\left(\frac{i}{N}\right)} (1 + o_p(1)) \end{aligned}$$

Entonces,

$$\begin{aligned} \Delta_N^1 &= (\Delta_N^{(1,1)}(\theta) + \Delta_N^{(1,2)}(\theta))(1 + o_p(1)) \\ E(\Delta_N^{(1,1)}(\theta)) &= \frac{-1}{n h_d h_r} \sum_{i=1}^N \int_0^1 K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{y - \frac{i}{N}}{h_r}\right) \frac{P(y) - P(\frac{i}{N})}{f(\frac{i}{N})} f(y) dy (1 + o(1)) = \\ &= \frac{-1}{h_d h_r} \int_0^1 \int_0^1 K_d\left(\frac{P(x) - \theta}{h_d}\right) K_r\left(\frac{y - x}{h_r}\right) \frac{P(y) - P(x)}{f(x)} f(y) dy dx (1 + o(1)) \end{aligned}$$

Como $|\frac{y-x}{h_r}| < 1$ tenemos que $x - h_r < y < x + h_r$

$$\int_{x-h_r}^{x+h_r} K_r\left(\frac{y-x}{h_r}\right) \frac{P(y) - P(x)}{f(x)} f(y) dy = \int_{-1}^1 K_r(v) h_r \frac{P(x+h_r v) - P(x)}{f(x)} f(x+h_r v) dv =$$

Donde en la última integral hemos hecho el cambio de variables $v = \frac{y-x}{h_r}$.

Ahora como $f(x + h_r v) = f(x) + f'(x) h_r v + o(h_r v)$, $P(x + h_r v) = P(x) + P'(x) h_r v + \frac{1}{2} P''(x) h_r^2 v^2 + o(h_r^2 v^2)$

$$= \int_{-1}^1 K_r(v) h_r \frac{P'(x) h_r v + \frac{1}{2} P''(x) h_r^2 v^2 + o(h_r^2 v^2)}{f(x)} (f(x) + f'(x) h_r v + o(h_r v)) dv =$$

$$\begin{aligned}
&= h_r^2 P'(x) \int_{-1}^1 K_r(v) v dv + \frac{1}{2} P''(x) h_r^3 \int_{-1}^1 v^2 K_r(v) dv + h_r^3 \frac{P'(x) f'(x)}{f(x)} \int_{-1}^1 K(v) v^2 dv + \\
&\quad + \frac{h_r^2}{2} \frac{P''(x) f'(x)}{f(x)} \int_{-1}^1 v^3 K_r(v) dv + o(h_r^3)
\end{aligned}$$

Entonces,

$$E(\Delta_N^{(1,1)}(\theta)) = -h_r^2 \kappa_2(K_r) \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-\theta}{h_d}\right) \left\{ P''(x) + \frac{2P'(x)f'(x)}{f(x)} \right\} \cdot (1 + o(1)) =$$

Haciendo el cambio de variable $z = P(x)$ y aplicando Bochner

$$= -h_r^2 \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) \cdot (1 + o(1))$$

Calculemos la varianza

$$Var(\Delta_N^{(1,1)}(\theta)) = \frac{N}{n^4 h_d^2 h_r^2} Var(\sum_{i=1}^N K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{P(\Theta_j) - P(\frac{i}{N})}{f(\frac{i}{N})}) \leq$$

usando el hecho que $E(\Theta^2) - E(\Theta)^2 \leq E(\Theta^2)$

$$\leq \frac{1}{n h_d^2 h_r^2} E\left(\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right)^2 (1 + o(1))\right) = o\left(\frac{1}{N h_d}\right).$$

Esto pues $|\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx| \leq C$, luego,

$$\begin{aligned}
&\frac{1}{N h_d^2 h_r^2} E\left(\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right)^2 (1 + o(1))\right) \leq \\
&C \frac{1}{n h_d^2 h_r^2} E\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right) (1 + o(1)) \\
&\leq \frac{1}{N h_d^2 h_r^2} C h_d h_r (h_r^2) \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) \\
&= \frac{h_r}{N h_d} C \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) = o\left(\frac{1}{N h_d}\right)
\end{aligned}$$

Esto implica operando y usando las condiciones asumidas, que

$$\Delta_N^{(1,1)}(\theta) + \kappa_2(K_r) h_r^2 \left(\frac{P''f + 2P'f'}{fP'} \right)(P^{-1}(\theta)) = o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

y por lo tanto 4.22 se deduce de

$$\sqrt{N h_d} \Delta_N^{(1,2)}(\theta) \xrightarrow{D} N(0, r^2(\theta))$$

Para probar esta afirmación notemos que $E[\Delta_N^{(1,2)}(\theta)] = 0$ y calculamos la varianza

$$\begin{aligned}
var(\sqrt{N h_d} \Delta_N^{(1,2)}(\theta)) &= \frac{1}{n^3 h_d h_r^2} \sum_{j=1}^N var \sum_{i=1}^N \left(K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{\sigma(\Theta_j) \epsilon_j}{f(\frac{i}{N})} \right) = \\
&= \frac{1}{h_d h_r^2} \int_0^1 \sigma^2(x) \int_0^1 \left(K_d\left(\frac{P(y)-\theta}{h_d}\right) K_r\left(\frac{x-y}{h_r}\right) \frac{dy}{f(y)} \right)^2 f(x) dx \cdot (1 + o(1)) = \\
&= \frac{1}{h_d h_r^2} \int_0^1 K_d\left(\frac{P(z)-\theta}{h_d}\right) \int_0^1 K_d\left(\frac{P(y)-\theta}{h_d}\right) \frac{1}{f(y)f(z)} \times \\
&\quad \times \int_0^1 \sigma^2(x) K_r\left(\frac{x-y}{h_r}\right) K_r\left(\frac{x-z}{h_r}\right) f(x) dx dy dz \cdot (1 + o(1)) =
\end{aligned}$$

Ahora hacemos el cambio de variables $w = \frac{P(z)-\theta}{h_d}$, entonces $z = P^{-1}(\theta + h_d w)$, $v = \frac{P(y)-\theta}{h_d}$, entonces $y = P^{-1}(\theta + h_d v)$, por último, $u = \frac{x-y}{h_r}$, entonces, $x = y + h_r u$ luego, $x = P^{-1}\theta + h - dv + h_r u$.

$$\frac{x-z}{h_r} = \frac{P^{-1}(\theta + h_d v) + h_r u - P^{-1}(\theta + h_d w)}{h_r} = \frac{P^{-1}(\theta + h_d v) - P^{-1}(\theta + h_d w)}{h_r} + u.$$

El Jacobiano nos da $|J| = h_r \left(\frac{h_d^2}{P(P^{-1}(\theta))^2} + o(h_d) \right)$ tenemos también

$P^{-1}(\theta + h_d w) = P^{-1}(\theta) + P^{-1'}(\theta) h_d w + P^{-1''} h_d^2 w^2$ entonces lo anterior queda:

$$\begin{aligned}
&= \frac{1}{h_d h_r^2} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta+h_d v))f(P^{-1}(\theta+h_d w))} \int_0^1 \sigma^2(P^{-1}(\theta+h_d v)+h_r u) K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) \frac{h_d^2 h_r}{P'(P^{-1}(\theta))^2} f(P^{-1}(\theta)+h_d+h_r u) du dv dw \\
&= \frac{h_d}{h_r} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta))f(P^{-1}(\theta))} \int_0^1 \sigma^2(P^{-1}(\theta)) K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) \frac{h_d^2 h_r}{P'(P^{-1}(\theta))^2} f(P^{-1}(\theta)) du dv dw = \\
&= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))^2 f(P^{-1}(\theta))} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta))f(P^{-1}(\theta))} \int_0^1 K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) du dv dw =
\end{aligned}$$

Haciendo el cambio de variables $y = \frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u$, $w = w$, $u = u$, $v = \frac{P(h_r(y-u))+P^{-1}(\theta+h_d u)-\theta}{h_d}$ cuyo Jacobiano $|J| = \frac{h_r}{h_d} P'(P^{-1}(\theta))(1+o(1))$

$$= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int \int \int K_d(w) K_d\left(\frac{P(h_r(y-u))+P^{-1}(\theta+h_d u)-\theta}{h_d}\right) K_r(u) \times K_r(y) du dw dy =$$

Aplicando Taylor tomando $x = P^{-1}(\theta+h_d w)$ tenemos que

$$\lim_{h_r \rightarrow 0, h_d \rightarrow 0} \frac{h_r}{h_d} K_d\left(\frac{P(P^{-1}(\theta+h_d v)+h_r(v-u))-\theta}{h_d}\right) = K_d(w + cP'(P^{-1}(\theta))(v-u))$$

, entonces,

$$= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int \int \int K_d(w) K_d(w + cP'(P^{-1}(\theta))(v-u)) K_r(u) \times K_r(y) du dw dy =$$

Con esto hemos probado la varianza asintótica en 4.17. Para la normalidad asintótica calculamos de manera análoga

$$\begin{aligned}
E((\sqrt{N} h_d \Delta_N^{(1,2)}(\theta))^4) &= \sum_{j=1}^N E \left[\left\{ \frac{\sigma(\Theta_j)}{n^{3/2} h_d^{1/2} h_r} \epsilon_j \sum_{i=1}^N K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{1}{f(\frac{i}{N})} \right\}^4 \right] = \\
&= \sum_{j=1}^N E \left[\int_0^1 \frac{(\sigma(\Theta) \epsilon_j)^4}{n^6 h_d^2 h_r^4} \left\{ n \int_0^1 K_d\left(\frac{P(y)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - y}{h_r}\right) \frac{1}{f(y)} dy \right\}^4 f(x) \right] = \\
&= \frac{n n^4}{n^6 h_d^2 h_r^4} E(\epsilon_j^4) \left[\int_0^1 \sigma^4(x) \left\{ \int_0^1 K_d\left(\frac{P(\Theta_1)-\theta}{h_d}\right) K_r\left(\frac{x-\Theta_1}{h_r}\right) \frac{1}{f(\Theta_1)} dx_1 \right\} f(x_1) \dots \right. \\
&\quad \left. \left\{ \int_0^1 K_d\left(\frac{P(\Theta_4)-\theta}{h_d}\right) K_r\left(\frac{x-\Theta_4}{h_r}\right) \frac{1}{f(\Theta_4)} dx_4 \right\} f(x_4) \right] =
\end{aligned}$$

Ahora hacemos el cambio de variable $y_1 = \frac{P(x_1)-\theta}{h_d}$ entonces $x_1 = P^{-1}(\theta+y_1 h_d)$, $y_2 = \frac{P(x_2)-\theta}{h_d}$
 $y_3 = \frac{P(x_3)-\theta}{h_d}$ $y_4 = \frac{P(x_4)-\theta}{h_d}$ $\tilde{x} = \frac{x-x_1}{h_r}$. El Jacobiano $|J| = \frac{h_d^4}{(P'(P^{-1}(\theta)))^5} h_r$

$$= \frac{N}{h_d^2 h_r^4} E(\epsilon_1)^4 \int_0^1 \sigma^4(P^{-1}(\theta)) \frac{h_d^4 h_r}{(f(P^{-1}(\theta)))^4 [P'(P^{-1}(\theta))]^5} \int_0^1 K_d(y_1) K_r(\tilde{x}) \cdot$$

$$\begin{aligned}
&\int K_d(y_2) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_1)-P^{-1}(\theta+h_d y_1)}{h_r}) \int_0^1 K_d(y_3) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_3)-P^{-1}(\theta+h_d y_3)}{h_r}) \\
&\int_0^1 K_d(y_4) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_4)-P^{-1}(\theta+h_d y_4)}{h_r}) =
\end{aligned}$$

Ahora $x = x_1 + h_r \tilde{x}$ entonces $x = P^{-1}(\theta+h_d y_1) + h_r \tilde{x}$

$$= \frac{\sigma^4(P^{-1}(\theta)) h_d^2}{N h_r^3} \frac{E(\epsilon_1)^4}{f(P^{-1}(\theta))^3 P'(P^{-1}(\theta))^5} \int \int K_d(y_1) K_r(\tilde{x}) \left[\prod_{j=2}^4 \int K_d(y_j) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_3)-P^{-1}(\theta+h_d y_3)}{h_r}) dy_j \right] dy_1 d\tilde{x}$$

Hacemos el cambio de variables $\tilde{x} = \tilde{x}$, $y_1 = y_1$ $y_j = \frac{P((P^{-1}(\theta+h_d \tilde{x})+h_r(y_j-y_1-z))}{h_d}$

calculemos $P^{-1}(\theta + h_d y_j) = \frac{P^{-1}(\theta + h_d P(P^{-1}(\theta + h_d \tilde{x}) + h_r(y_j - y_1) - \theta))}{h_d} = P^{-1}(\theta + h_d \tilde{x}) + h_r(y_j - y_1)$.

Calculemos el Jacobiano $|J| = \frac{h_r^3 P'(P^{-1}(\theta))^3}{h_d^3}$

Entonces,

$$= \frac{\sigma^4(P^{-1}(\theta))h_d^2}{N h_r^3} \frac{E(\epsilon_1)^4}{f(P^{-1}(\theta))^3 P'(P^{-1})^5} \int \int K_d(y_1) K_r(\tilde{x}).$$

$$\left[\prod_{j=2}^4 \int K_r(y_j) K_d(y_1 + \frac{h_r}{h_d} P'(P^{-1}(\theta)(y_j - \tilde{x})) dy_j \right] d\tilde{x} dy_1 \cdot \frac{h_r^3 P'(P^{-1}(\theta))^3}{h_d^3} = O(\frac{1}{N h_d}) = o(1)$$

Teorema 4.2.13. Asuma las hipótesis del teorema anterior y sean \hat{P}_I y P_T denotan las inversas de las funciones \hat{P}_I^{-1} y P_T^{-1} definidas por 4.7 y 4.8 respectivamente. Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$ existe, entonces tenemos, que para $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{N h_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{f} \right) (\theta) \right) \Rightarrow N(0, s^2(\theta))$$

donde la varianza asintótica está dada por

$$s^2(\theta) = \frac{\sigma^2(\theta) P'(\theta)}{f(\theta)} \int \int \int K_d(w + c P'(\theta)(v - u)) K_d(w) K_r(u) K_r(v) dw du dv$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$ tenemos que para todo $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{N h_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{f} \right) (\theta) \right) \Rightarrow N(0, \tilde{s}^2(\theta))$$

donde la varianza asintótica está dada por $\tilde{s}^2(\theta) = \frac{\sigma^2(\theta)}{f(\theta)} \int K_r^2(u) du$.

Demostración. Del Lema 4.2.10 obtenemos del desarrollo de Taylor

$$H_2^{-1}(\theta) - H_1^{-1} = \Phi(H_2)(\theta) - \Phi(H_1)(\theta) = \Phi(H_1 + H_2 - H_1) - \Phi(H_1) = Q(1) - Q(0) = Q'(0) + \frac{1}{2} Q''(\lambda^*)$$

$$\text{con } H_2 = \hat{P}_I^{-1} \quad H_1 = P_N^{-1}.$$

Definamos además $A_N = Q'(0) = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta)$ y

$$B_N = Q''(\lambda^*) = \frac{2(\hat{P}_I^{-1} - P_N^{-1})(\hat{P}_I^{-1} - P_N^{-1})'}{P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1})} \circ (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))(\theta) -$$

$$- \frac{(\hat{P}_I^{-1} - P_N^{-1})^2 (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))''}{(P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^3} \circ (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))(\theta)$$

o sea $P_I(\theta) - P_N(\theta) = A_N + \frac{1}{2} B_N$. Mostraremos que

$$(4.23) \quad A_N = -\frac{(\hat{P}_I^{-1} - P_N^{-1})}{(P^{-1})'} \circ P(\theta) + o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

y que,

$$(4.24) \quad B_N = o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

Para probar 4.23 consideramos

$$D_N = (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P(\theta) = (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta))'(\zeta_N)(P_N(\theta) - P(\theta))$$

donde $|\zeta_N - P(\theta)| \leq |P_N(\theta) - P(\theta)|$. Como $(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) = \Delta_N$, entonces,

$$(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta))'(\zeta_N) = \Delta'_N(\zeta_N) = \frac{1}{N h_d} \sum_{i=1}^N (K_d(\frac{\zeta_N - \hat{P}(\frac{i}{N})}{h_d}) - K_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})) =$$

$$= \frac{1}{Nh^2} \sum_{i=1}^N (K'_d(\frac{\zeta_{i,n}-\varphi_N}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N})))$$

$$\text{con } |\zeta_{i,n} - P(\frac{i}{N})| \leq |\hat{P}(\frac{i}{N}) - P(\frac{i}{N})| = O(R_N) \text{ c.s. siendo } R_N = \left(\frac{\log h_r^{-1}}{Nh_r}\right)^{1/2}$$

$$\Delta'_N(\zeta_N) = \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N}))) + \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\eta - \varphi_N}{h_d}) - K'_d(\frac{P(\frac{i}{N}) - \varphi_N}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N})))$$

donde para el segundo término de la última igualdad tenemos que

$$L_N = \frac{N}{Nh_d^2} K''_d(\frac{u - \varphi_N}{h_d})(\frac{\eta - \varphi_N}{h_d} - \frac{P(\frac{i}{N}) - \varphi_N}{h_d}) O(R_N) = \frac{O(R_N^2)}{h_d^3}$$

Entonces tenemos que:

$$\begin{aligned} \Delta'_N(\zeta_N) &= \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N}))) + O(\frac{R_N^2}{h_d^3}) \text{ c.s.} \\ &= \frac{1}{h_d^2} \int (K'_d(\frac{P(x) - P(\theta)}{h_d})(\hat{P}(x) - P(x))) dx + O(R_N + \frac{R_N^2}{h_d^3} + \frac{1}{Nh_d}) \text{ c.s.} \\ &= O(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^2} + \frac{1}{Nh_d}) \text{ c.s.} \end{aligned}$$

Tenemos que: $D_N = (\hat{P}_I^{-1} - P_N^{-1})(\varphi_N)(P_N(\theta) - P(\theta)) = O(R_N h_d + \frac{R_N^2}{h_d} + \frac{h_d}{N}) = o(\frac{1}{\sqrt{Nh_d}})$ c.s. luego como $(P_N^{-1})'(\theta) = (P^{-1})'(\theta) + o(1)$ sale la igualdad

$$A_N = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta) = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P(\theta) + D_N = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta) + o(\frac{1}{\sqrt{Nh_d}}) \text{ c.s.}$$

Para B_N notemos que $B_N = 2B_{N1} - B_{N2}$ con

$$B_{N1} = \frac{(\hat{P}_I^{-1} - P_N^{-1})(\hat{P}_I^{-1} - P_N^{-1})'(\theta_N)}{(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^2(\theta_N)} \text{ y } B_{N2} = \frac{(\hat{P}_I^{-1} - P_N^{-1})^2(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))''}{(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^3(\theta_N)}$$

donde $\theta_N = (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^{-1}(\theta)$. Como $\hat{P}_I^{-1} - P_N^{-1} \rightarrow 0$ y $P_N^{-1} \rightarrow P^{-1}$ entonces $\theta_N \rightarrow (P^{-1})^{-1}(\theta) = P(\theta)$ de donde el denominador de B_{N1} tiende a θ^2 .

Recordemos que si $\sqrt{Nh_d}Z \rightarrow N(0, \sigma^2)$ entonces $Z = O_p(\frac{1}{\sqrt{Nh_d}})$ y si $\Theta_N + a_N = O_p(b_N)$ entonces $\Theta_N = O_p(b_N)$ luego $P_I^{-1} - P_N^{-1} = O_p(\frac{1}{\sqrt{Nh_d}})$ además $(P_I^{-1} - P_N^{-1})' = O_p(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^3} + \frac{1}{Nh_d}) = O_p(\frac{1}{\sqrt{Nh_d}}(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^3})) = o_p(\frac{1}{\sqrt{Nh_d}})$ pues $\frac{R_N}{h_d} \rightarrow 0$ ya que $\frac{(\log h_r^{-1})^{1/2}}{n^{1/2}h_r^{1/2}h_d} = \frac{(\log h_r^{-1})^{1/2}h_r h_r}{n^{1/2}h_r^{5/2}h_d} \rightarrow 0$ y $\frac{R_N^2}{h_d^3} = \frac{\log h_r^{-1}}{Nh_r h_d^3} = o(1)$ por hipótesis o sea $B_{n1} = o_p(\frac{1}{\sqrt{Nh_d}})$

Igualmente $B_{n2} = O_p(\frac{1}{Nh_d})$ lo que prueba la segunda igualdad.

Tenemos que:

$$\begin{aligned} &\sqrt{Nh_d}(\hat{P}_I(\theta) - P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= \sqrt{Nh_d}(\frac{\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)}{(P_N^{-1})'} \circ P_N(\theta) + \frac{1}{2}B_N - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= -\sqrt{Nh_d}(\frac{\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)}{(P_N^{-1})'} \circ P_N(\theta) + o_p(\frac{1}{\sqrt{Nh_d}}) + \frac{1}{2}o_p(\frac{1}{\sqrt{Nh_d}}) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= -\frac{\sqrt{Nh_d}(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)(P_N^{-1})' \circ P_N(\theta)}{(P^{-1})' \circ P(\theta)} + o_p(1) = \\ &= -P'(\theta)\sqrt{Nh_d}(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{P'f})(\theta)) + o_p(1) = Z_N \end{aligned}$$

Usando el teorema anterior en $P(\theta)$ $Z_N \Rightarrow N(0, s^2(\theta))$ con

$$\begin{aligned} s^2(\theta) &= \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \times \int \int \int K_d(w + cP'(P^{-1}(\theta)))(v - u)K_d(w)K_r(u)K_r(v)dwdudv = \\ &= \frac{\sigma^2(\theta)P'(\theta)}{f(\theta)} \times \int \int \int K_d(w + cP'(P^{-1}(\theta)))(v - u)K_d(w)K_r(u)K_r(v)dwdudv \end{aligned}$$

lo que demuestra el teorema.

Capítulo 5

Simulaciones

Como hemos visto en los capítulos precedentes existen múltiples condiciones que influyen en la performance de los estimadores, a saber, el tipo de núcleo, el ancho de la ventana, la cantidad de ítems que integran el test, el número de sujetos a los que se aplica el test, la dimensión y la estructura de dependencia entre los rasgos. En este capítulo presentamos una serie de estudios basados en simulación para estudiar la incidencia de estas condiciones en la estimación, tanto del rasgo como de las CCI.

Se realizaron 17 estudios, que no abarcan todos los posibles cruzamientos de los factores que inciden en la performance de los estimadores. Simplemente queremos atacar ciertos aspectos que en el desarrollo teórico quedaron sin resolver. También queremos estudiar la robustez de los estimadores ante determinados componentes como por ejemplo la elección del núcleo y la distribución de los rasgos.

Estudio 1: Ajuste en la estimación del rasgo.

Este estudio tiene como objetivo estudiar el ajuste de la estimación del rasgo mediante el modelo no paramétrico. Para ello se realizó un análisis unidimensional y luego un análisis bidimensional.

En el análisis unidimensional se comparó el ajuste de la estimación no paramétrica del rasgo mediante el método propuesto y mediante la estimación paramétrica correspondiente a cada modelo. Para calcular el ajuste se utilizó una distancia (Dif) la cual está definida en la sección 5.1.1. También se calculó la correlación entre el rasgo simulado y las estimaciones no paramétricas y paramétricas.

En el análisis bidimensional se calculó la correlación entre la estimación del rasgo en cada dimensión y el real. También se comparó la correlación entre los rasgos estimados y entre los rasgos reales (la estructura de dependencia) y la norma de las diferencias en R^2 mediante la medida Dif.

Estudio 2: Distribución de la estimación del rasgo.

Un aspecto que la tesis no contiene es la deducción de la distribución exacta o asintótica del estimador del rasgo. Existen en la literatura aproximaciones de la distribución del puntaje total de un sujeto que responde ítems con distinta probabilidad de acierto vía la distribución compuesta binomial o la Poisson Binomial. En nuestro caso el problema es diferente y puede enunciarse así: si los sujetos puntúan en ciertos estadísticos (eventualmente en el puntaje total) y se toma como estimador del rasgo del sujeto k el cuartil empírico de esos puntajes, ¿cuál es la distribución de este estimador? Sabemos que el estimador converge casi seguramente al valor del rasgo, pero no hemos podido calcular su sesgo ni varianza y tampoco hemos podido aplicar el teorema central del límite para ver si asintóticamente se distribuye normalmente. Es así que

recurriremos a la simulación para poder analizar este punto.

Estudio 3: Efecto del tamaño muestral en la estimación de las CCI unidimensionales. En este estudio se muestra la relación entre el tamaño muestral y la exactitud en la estimación de las CCI.

Estudio 4: Tamaño muestral en dos dimensiones. Caso I. Se simularán ítems que dependen de un solo rasgo.

Estudio 5: Tamaño muestral en dos dimensiones. Caso II. Se simularán ítems que dependen de los dos rasgos.

Estudio 6: Ajuste por ítems y sujetos en dos dimensiones. Para dos dimensiones se analizó el ajuste en función del tamaño muestral y el número de ítems.

Estudio 7: Efecto de la elección de la ventana en una dimensión. Se comparó el ajuste para distintos valores de la ventana.

Estudio 8: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso I.

Estudio 9: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso II.

En el estudio 8 y 9 se eligieron diferentes matrices de anchos de banda y se comparó su ajuste en el caso de ítems dependientes de un solo rasgo e ítems dependientes de dos rasgos.

Estudio 10: Efecto del tipo de núcleo en una dimensión.

Estudio 11: Efecto de distintos núcleos en dos dimensiones. Caso I.

Estudio 12: Efecto de distintos núcleos en dos dimensiones. Caso II.

En los estudios 10, 11 y 12 se analizó el efecto en el ajuste de la elección del núcleo. En el estudio 10 se utilizan núcleos unidimensionales y en los estudios 11 y 12 se usan núcleos bidimensionales, distinguiendo como antes el caso I y II antes descrito.

Estudio 13: Efectos de la distribución de los rasgos en una dimensión.

Estudio 14: Efecto de distintas distribuciones de θ en dos dimensiones. Caso I.

Estudio 15: Efecto de distintas distribuciones de θ en dos dimensiones. Caso II.

Estudio 16: Efecto de la elección de la ventana en el modelo isótono.

Estudio 17: Efecto de la elección del núcleo en el modelo isótono.

Las simulaciones se realizaron con el programa R y se usaron los paquetes Psych, Copula, MASS, R2Cuba, Truncnorm, Sn, Pracma, Ltm, Mirt, Mokken, KernSmoothIRT, Stats, Multcomp, GPArotation, Ks y Rgl. Además, se generaron funciones específicas para la simulación y estimación de modelos unidimensionales y multidimensionales.

5.1. Método

Esta sección está destinada a presentar los componentes comunes a todos los estudios. Los elementos específicos se presentarán en la sección respectiva de cada estudio. Se tratarán tres aspectos:

- a) Medidas de ajuste para el rasgo y para las CCI.
- b) Modelos de TRI utilizados en las simulaciones.
- c) Generación de los rasgos y de los parámetros de los ítems.

Esperamos al incluir este tema al inicio del capítulo ayudar a entender mejor lo que se presenta y también ayudar a quien quiera replicar los estudios.

Para analizar el ajuste de los rasgos estimados al rasgo real y de las curvas o superficies de respuesta estimadas a las respectivas simuladas se definieron diversas medidas de ajuste, que se describen en las siguientes dos subsecciones.

5.1.1. Medidas de ajuste de los rasgos

Para estimar la exactitud de la estimación de los rasgos se utilizó una medida de ajuste basada en la norma de las diferencias. Se consideró el promedio de las normas de las diferencias.

Específicamenete se definió la medida Dif mediante:

$$Dif = N^{-1} \sum_{i=1}^N \| \theta_i - \hat{\theta}_i \|$$

5.1.2. Distancia entre curvas características del ítem

A los efectos de estimar el ajuste de las curvas trabajamos con varias distancias entre las curvas características de los ítems o de las superficies de respuesta. Con la finalidad de simplificar, y siempre que no exista lugar a confusión, se utilizará el término CCI, aunque estemos en dimensiones mayores a 1. También, se emplearán las siglas en español, aunque sean conocidas generalmente por sus siglas en inglés.

Sea \mathcal{A} la familia de las CCI.

Definición 5.1.1. Definimos las distancias $d_i : \mathcal{A} \times \mathcal{A} \rightarrow R$ con $i=1,2,3,4$ como:

Raíz del error cuadrático integrado (RECI):

$$RECI = d_1(P_1, P_2) = \left\{ \int \cdots \int_{R^d} |P_1(\theta) - P_2(\theta)|^2 d\theta \right\}^{1/2}$$

Si tomamos la esperanza del error cuadrático obtenemos d_2 o raíz del error cuadrático esperado (RECE).

$$RECE = d_2(P_1, P_2) = \left\{ \int \cdots \int_{R^d} |P_1(\theta) - P_2(\theta)|^2 f(\theta) d\theta \right\}^{1/2}$$

Definimos la tercera distancia como la distancia del supremo:

$$d_3(P_1, P_2) = \sup_{\theta} |P_1(\theta) - P_2(\theta)|$$

Observación: la distancia del supremo está asociada a la convergencia uniforme de las sucesiones de funciones.

Desviación absoluta media (DAM):

$$DAM = d_4(P_1, P_2) = \int \cdots \int_{R^d} |P_1(\theta) - P_2(\theta)| f(\theta) d\theta$$

5.1.3. Modelos de TRI utilizados

En las simulaciones utilizaremos diversos modelos de TRI que representan los modelos típicos que encontramos en la bibliografía. Nos restringiremos a los modelos para ítems dicotómicos pues son los que analizamos en profundidad en esta tesis. En dimensión 1 se utilizaron modelos logísticos de 1, 2 y 3 parámetros, y los modelos de ojiva normal de 1, 2 y 3 parámetros, ya que son los que clásicamente se utilizan. Para ejemplificar modelos que no sean necesariamente crecientes y que no ajusten con los métodos usuales de estimación usaremos modelos logit cúbicos, ya que estos tienen puntos de inflexión. Si se toman b y c positivos, estos modelos son crecientes.

Los modelos son:

1. $P(\theta) = \frac{1}{1+e^{-(\theta-b)}}$
2. $P(\theta) = \frac{1}{1+e^{-a(\theta-b)}}$
3. $P(\theta) = c + (1-c) \frac{1}{1+e^{-a(\theta-b)}}$
4. $P(\theta) = \int_{-\infty}^{(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$
5. $P(\theta) = \int_{-\infty}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$
6. $P(\theta) = c + (1-c) \int_{-\infty}^{a(\theta-b)} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$
7. $\log \frac{P(\theta)}{1-P(\theta)} = a + b\theta + c\theta^3$

Para las simulaciones de dimensión superior se utilizará el modelo de ojiva normal multivariado, el modelo logístico multivariado y el cúbico multidimensional.

Sean las variables dicotómicas $\nu_i = \{0, 1\}$, y la función que modela la probabilidad de respuesta es la función de distribución de la normal estándar, $\Phi(\bullet)$, las funciones de respuesta vienen dadas por:

$$\Pr(X_i = 1 | \Theta = \theta) = \Phi(\alpha_i + \beta_i' \theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha_i + \beta_i' \theta} \exp\left(-\frac{y^2}{2}\right) dy$$

También se utilizó el modelo logístico multidimensional:

$$P(U_{ij} = 1 | a_i, d_i, c_i, \theta_j) = c_i + (1 - c_i) \frac{e^{(a_i^t \theta_j + d_i)}}{1 + e^{(a_i^t \theta_j + d_i)}}$$

donde t indica el vector traspuesto.

Como modelo alternativo a los usuales se usó el logit cúbico multidimensional que toma la forma:

$$\log \frac{P(\theta)}{1-P(\theta)} = a + b^t \theta + c^t \theta^3$$

5.1.4. Generación de datos

Para generar tests simulamos los rasgos de los examinados, así como los valores de otros parámetros relacionados con los distintos modelos. Se nombró cada tipo de simulación de acuerdo al siguiente esquema: las primeras dos letras indican el modelo (P1, P2, P3 (logísticos de 1,2 y 3 parámetros), O1, O2, O3 (ojiva normal) y L3 (logit cúbico)). Luego, se indicó la cantidad de sujetos y por último, la cantidad de ítems utilizados. Así, por ejemplo, P3N500n50 indica que la simulación se basó en el modelo 3P, se simularon 500 sujetos que respondieron a 50 ítems. No agregamos la dimensión en la descripción pues se realizan en instancias separadas y no da lugar a confusión.

En el caso de los modelos unidimensionales los parámetros a , b y c los simulamos a partir de variables aleatorias uniformes, y sus distribuciones fueron: $a \sim U(0, 5; 2)$, $b \sim U(-3; 3)$ y $c \sim U(0; 0, 25)$. Para el modelo logit cúbico unidimensional $a \sim U(-1; 1)$, $b \sim U(-2; 0.5)$ y $c \sim U(0.5; 1.5)$.

En el caso de los modelos multidimensionales para el modelo de ojiva normal se consideró $c \sim U(0; 0, 25)$; $\alpha_i \sim U(-3; 3)$; y cada componente de β_i se consideraron con distribución $U(0, 5; 2)$. Para el modelo logístico multidimensional se utilizó $c \sim U(0; 0, 25)$; $d_i \sim U(-3; 3)$; y cada componente de $a_i \sim U(0, 5; 2)$. En el modelo logit cúbico multidimensional $a \sim U(-1; 1)$,

y los vectores $b \sim U(-2; 0.5)$ y $c \sim U(0.5; 1.5)$. Cuando se usaron rasgos en un soporte acotado los parámetros de los ítems se transformaron para adecuarse a dicho soporte.

Para simular los rasgos (θ) se usaron algunas distribuciones que pueden caracterizar distintos casos de los rasgos. Las distribuciones usadas fueron:

Distribución Normal $N(\mu, \sigma^2)$.

La densidad viene dada por $f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ donde el valor esperado es μ y la varianza σ^2 .

Distribución beta. Es una distribución con dominio en $[0, 1]$ y puede ser usada para representar rasgos donde se relaciona las probabilidades con la proporción de aciertos.

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}$$

donde Γ es la función gamma. Se tiene que $E[X] = \frac{a}{a+b}$ y $V[X] = \frac{ab}{(a+b+1)(a+b)^2}$.

Distribución uniforme en $[0, 1]$.

Es un caso particular de la distribución beta con $a=1$ y $b=1$. Su función de densidad es

$$f(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x < 0 \text{ or } x > 1 \end{cases}$$

la media es $\frac{1}{2}$ y la varianza $\frac{1}{12}$.

Distribución arcoseno en $[0, 1]$.

Es un caso particular de la distribución beta con $a=b=1/2$. Su densidad es $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ con $E(X) = \frac{1}{2}$ y $V(X) = \frac{1}{8}$.

Distribución normal truncada.

Sea $Y \sim N(\mu, \sigma^2)$. Diremos que X tiene distribución normal truncada si tienen la distribución de Y condicionada a que $Y \in (a, b)$, $-\infty \leq a < b \leq \infty$.

La densidad es para $a \leq x \leq b$

$$f(x; \mu, \sigma, a, b) = \frac{\frac{1}{\sigma} \phi(\frac{x-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}$$

y 0 en otro caso.

La función $\Phi(\cdot)$ indica la función de distribución de la normal estándar.

$$E(X) = \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} =$$

$$V(X) = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma} \phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma} \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \right)^2 \right]$$

Esta distribución es usada en el modelo probit y tobit para respuestas binarias.

5.2. Estudio 1: Ajuste en la estimación del rasgo.

En esta sección se estudia el ajuste del rasgo estimado, para ello se utilizó la medida Dif propuesta anteriormente. Se analizó el efecto de la cantidad de ítems y sujetos utilizados, así como la distribución supuesta del rasgo. Primeramente, se presenta qué sucede en el caso unidimensional y luego el caso bidimensional. Se analizan tres aspectos que pueden influir en la estimación del rasgo, a saber, el tipo de modelo, la cantidad de ítems y la cantidad de sujetos. Se utilizaron para una dimensión los 7 tipos de modelos ya mencionados, la cantidad de ítems a simular fueron 10, 20, 50, 100 y 150; y la cantidad de sujetos a simular fueron 100, 250, 500, 750, 1000, 2000 y 5000.

Se realizaron 245 simulaciones por lo que se tuvieron para cada tipo de modelo 35 simulaciones, para cada cantidad de ítems 49 y para cada condición sobre el número de sujetos 35 simulaciones.

Para determinar el ajuste usamos la medida Dif y la correlación de la estimación del rasgo estimado con el verdadero cuando se utiliza estimación no paramétrica y cuando se usa la estimación paramétrica usual.

Resultados

En los casos unidimensionales, a partir de las simulaciones, se calcularon la distancia Dif y las correlaciones entre el rasgo estimado mediante el modelo no paramétrico y el real (simulado). También se calculó la correlación entre el rasgo estimado paramétricamente según el modelo adecuado y el real. Luego, se estimó la media de estas medidas. Cuando se analizan los resultados unidimensionales encontramos que la medida Dif es pequeña. Cuando observamos las medias de las correlaciones encontramos que, tanto la correlación del modelo paramétrico como no paramétrico, son altas y de magnitud similar.

La Dif media es menor en todos los casos cuando se estima el rasgo mediante el método no paramétrico, aunque la diferencia no llega a ser significativa.

También se verifica que el mayor error se da en el modelo logit cúbico, si bien igualmente es pequeño. Sería esperable que en el caso no paramétrico la estimación del rasgo debería comportarse igual de bien en el modelo logit cúbico que en los otros modelos (no así cuando se realiza una estimación paramétrica). Sospechamos que esto sucede porque se utiliza como estadístico para ordenar a los sujetos la suma de los aciertos y que si utilizáramos una suma adecuadamente ponderada mejoraría el ajuste en el modelo logit cúbico.

Las correlaciones de los estimados con el rasgo real cuando consideramos estimación paramétrica varían entre 0.91 a 0.98 y cuando usamos estimación no paramétrica la correlación varía entre 0.85 a 0.97. Se puede comprobar cómo el error disminuye al aumentar el número de ítems. En tanto, el número de sujetos no influye por sí solo en el ajuste del rasgo.

Concretamente podemos afirmar que la estimación no paramétrica es del orden de la paramétrica siendo mucho más simple de obtener.

A partir de lo encontrado, concluimos que el método no paramétrico propuesto de estimación del rasgo es muy adecuado y funciona igual que el paramétrico, siendo mucho más sencillo de obtener.

La tabla que sigue contiene los resultados obtenidos en el caso unidimensional. La primera columna muestra las variables consideradas, la segunda las medias del ajuste Dif cuando el rasgo es estimado según el método no paramétrico. La tercera columna contiene las medias del ajuste Dif halladas según el método paramétrico. La cuarta y quinta columnas tienen las medias de las correlaciones entre el rasgo simulado y el estimado no paramétricamente y paramétricamente respectivamente.

variable	media Dif NP	media Dif. PA	media corr..NP	media.corr.PA.
Modelo P1	0.29 (0.14)	0.30 (0.15)	0.96	0.97
Modelo P2	0.28 (0.13)	0.36 (0.16)	0.96	0.97
Modelo P3	0.34 (0.16)	0.41 (0.17)	0.95	0.96
Modelo O1	0.28 (0.13)	0.30 (0.14)	0.96	0.97
Modelo O2	0.27 (0.13)	0.35 (0.16)	0.97	0.97
Modelo O3	0.35 (0.16)	0.42 (0.16)	0.94	0.96
Modelo L3	0.62 (0.08)	0.63 (0.11)	0.85	0.91
10 ítems	0.56 (0.09)	0.60 (0.08)	0.88	0.91
20 ítems	0.42 (0.11)	0.51 (0.12)	0.93	0.95
50 ítems	0.31 (0.13)	0.38 (0.13)	0.96	0.97
100 ítems	0.24 (0.14)	0.27 (0.12)	0.97	0.98
150 ítems	0.21 (0.14)	0.22 (0.12)	0.97	0.98
100 sujetos	0.35 (0.19)	0.43 (0.20)	0.94	0.95
250 sujetos	0.36 (0.18)	0.40 (0.18)	0.94	0.95
500 sujetos	0.36 (0.17)	0.40 (0.18)	0.94	0.96
750 sujetos	0.34 (0.18)	0.39 (0.17)	0.94	0.96
1000 sujetos	0.33 (0.17)	0.38 (0.18)	0.95	0.96
2000 sujetos	0.34 (0.18)	0.39 (0.19)	0.94	0.96
5000 sujetos	0.34 (0.19)	0.38 (0.19)	0.94	0.96

Cuadro 5.1: Ajuste de los rasgos

Los siguientes gráficos muestran los datos relevantes, en rojo se tiene la media del ajuste Dif para la estimación no paramétrica del rasgo y en azul, la media del ajuste Dif para la estimación paramétrica del rasgo.

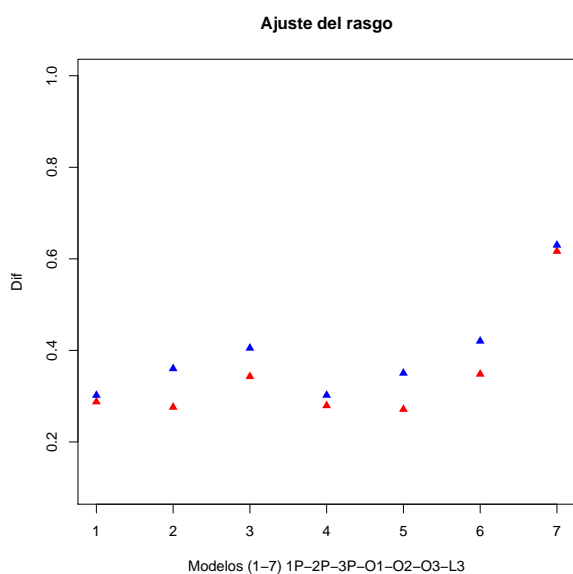


Figura 5.1: Ajuste por modelo

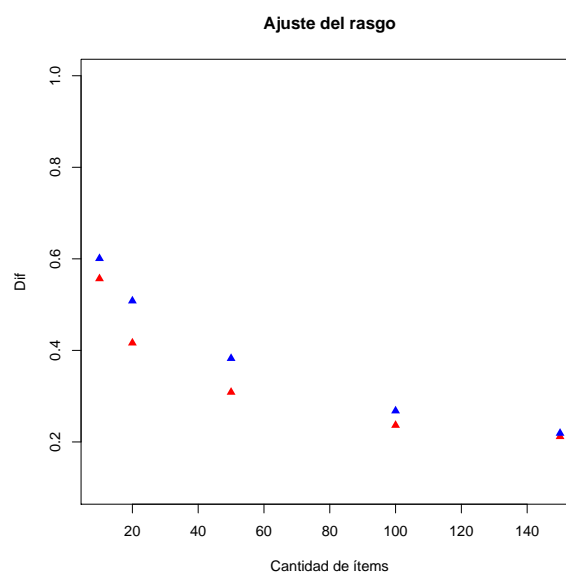


Figura 5.2: Ajuste por cantidad de sujetos

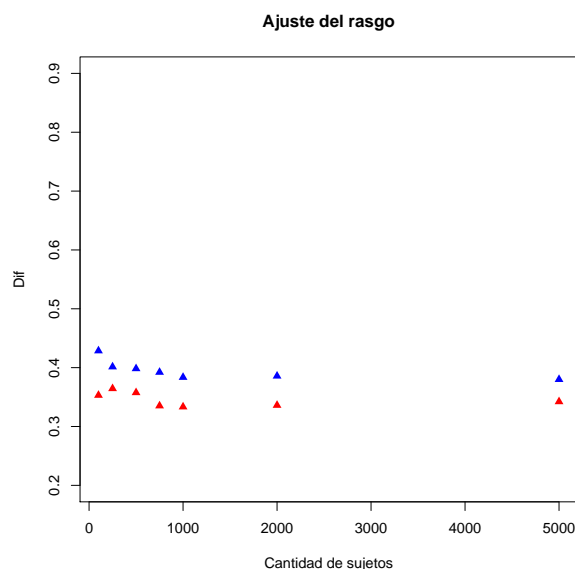


Figura 5.3: Ajuste por cantidad de sujetos

No podemos mostrar los resultados de todas las simulaciones, por lo extenso que resultaría. Por eso, en la tabla que sigue se muestran algunos de los resultados de las simulaciones realizadas para remarcar los aspectos más importantes.

Se observa claramente la relación entre el tamaño muestral y la cantidad de ítems, que refleja lo probado en la consistencia conjunta. Cuando hay poca cantidad de ítems, no importa cuanto aumente el tamaño muestral, el error es mayor. Tampoco disminuirá el error si aumentan los ítems, pero no la cantidad de sujetos. Adicionalmente, vemos un aspecto no tan esperado que es el aumento del error en los modelos logit cúbicos, ya que el proceso de estimación es aparentemente independiente del modelo. De esto último dimos una explicación anteriormente.

Estima	Dif NP	Dif PA	corr.NP	corr.PA
P1N100n10	0.49	0.53	0.91	0.93
P1N1000n10	0.51	0.54	0.90	0.92
P1N100n50	0.25	0.24	0.98	0.98
P1N750n50	0.24	0.26	0.98	0.98
P1N1000n100	0.16	0.17	0.99	0.99
P1N1000n150	0.13	0.13	0.99	0.99
P1N2000n150	0.14	0.15	0.99	0.99
P2N250n10	0.50	0.53	0.90	0.92
P2N2000n10	0.53	0.59	0.90	0.92
P2N100n20	0.36	0.62	0.95	0.96
P2N250n20	0.31	0.50	0.96	0.96
P2N2000n50	0.22	0.34	0.98	0.98
P2N1000n100	0.16	0.29	0.99	0.99
P2N2000n100	0.16	0.18	0.99	0.99
P3N100n10	0.64	0.72	0.85	0.89
P3N2000n10	0.63	0.65	0.85	0.88
P3N2000n50	0.28	0.38	0.97	0.98
P3N1000n150	0.17	0.18	0.99	0.99
P3N5000n150	0.17	0.18	0.99	0.99

L3N750n100	0.57	0.52	0.86	0.93
L3N1000n100	0.53	0.52	0.88	0.93
L3N2000n100	0.59	0.54	0.86	0.92
L3N5000n150	0.53	0.49	0.88	0.93

Cuadro 5.2: Simulaciones seleccionadas

Para el caso bidimensional, se procedió a simular test de 76,100 y 140 ítems para 500, 750 y 1000 sujetos. La correlación entre las dimensiones de los rasgos se simuló mediante un número aleatorio con distribución uniforme entre $[0.55 \text{ y } 0.75]$. El cuadro que sigue resume la información obtenida. En las primeras dos columnas encontramos las características de la simulación. En la columna tres y cuatro se recoge la correlación entre el rasgo simulado en cada dimensión y el estimado no paramétricamente. En la columna 5 se tiene la correlación entre las dimensiones simuladas y en la columna 6 la correlación entre las dimensiones estimadas. Por último, en la columna 7, se encuentra la medida Dif.

Sujetos	Ítems	Corr dim1	Corr dim2	Corr. dim reales	Corr dim estim.	Dif
500	76	0.81	0.81	0.62	0.54	0.04
500	100	0.82	0.83	0.62	0.55	0.04
500	140	0.86	0.87	0.56	0.55	0.04
750	76	0.81	0.79	0.70	0.62	0.04
750	100	0.82	0.80	0.65	0.62	0.03
750	140	0.83	0.82	0.63	0.57	0.03
1000	76	0.76	0.76	0.73	0.62	0.03
1000	100	0.78	0.78	0.67	0.63	0.03
1000	140	0.85	0.83	0.65	0.61	0.03

Cuadro 5.3: Ajuste en dos dimensiones

Cuando analizamos los modelos de dos dimensiones encontramos que:

- Las correlaciones de los rasgos estimados en cada dimensión con el rasgo verdadero correspondiente es alta.
- Se verifica que las correlaciones entre las dimensiones estimadas es aproximadamente igual a la correlación entre las dimensiones de los rasgos verdaderos, es decir, la estructura de dependencia se estima correctamente.
- La norma de las Dif es pequeña.
- La correlación entre los rasgos estimados es menor a la correlación entre los rasgos simulados en todos los casos, aunque la diferencia no es muy grande.

En este caso también se comprueba la buena aproximación que tiene el método propuesto.

Estimación de la cópula y la densidad conjunta

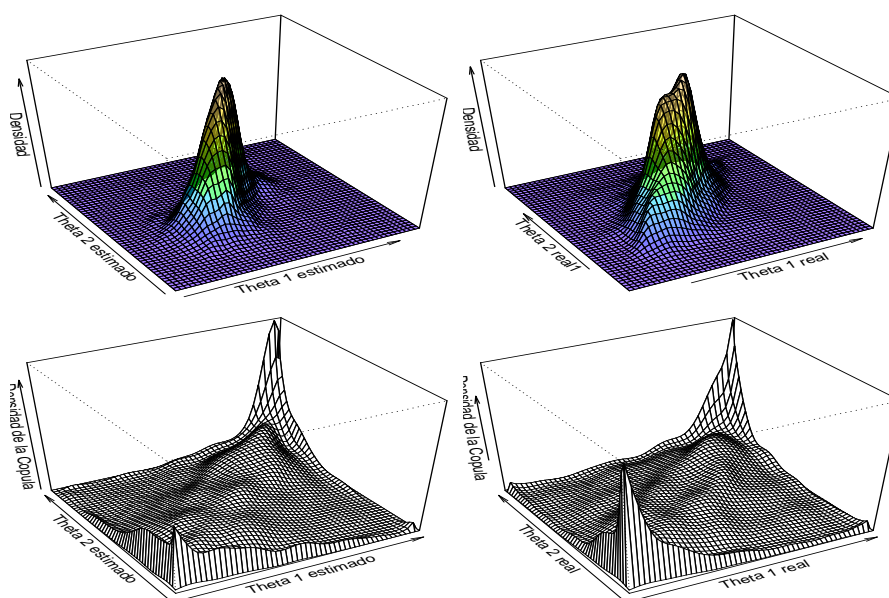


Figura 5.4: Densidad y cópula del rasgo

En dos dimensiones podemos estimar la densidad conjunta y la cópula. La gráfica anterior nos muestra la densidad estimada y real y la cópula estimada y real. Este caso corresponde a 1000 sujetos, 140 ítems y una correlación entre rasgos de 0.65.

5.3. Estudio 2: Distribución de la estimación del rasgo

Si bien hemos podido demostrar la consistencia conjunta de los rasgos y las CCI, no se pudo encontrar la distribución asintótica de los estimadores de los rasgos. Este es un elemento importante no resuelto aún y que, por ejemplo, en Ramsay (1991) se consideran los rasgos estimados sin error para poder determinar intervalos de confianza para las CCI. En este apartado analizamos la varianza de los estimadores mediante simulaciones. Hay varios elementos a tener en cuenta: en primer lugar, en la estimación interviene la distribución empírica por lo que nos restringimos a la distribución uniforme en $[0,1]$. Es claro que lo que interviene en la varianza de este rasgo es la cantidad de sujetos y la cantidad de ítems. A partir de esto simulamos test dicotómicos con 10, 25, 50, 75 y 100 ítems y aplicaciones a 100, 250, 500, 750 y 1000 sujetos con rasgo con distribución uniforme en $[0,1]$ y se agregaron 51 sujetos con rasgos igualmente espaciadas en el intervalo unidad (la distancia de los rasgos de los sujetos es 0.02). Estos sujetos se agregaron con rasgo fijo para así encontrar la distribución del rasgo usando replicaciones. Los test siguen un modelo logístico 3P con parámetros simulados aleatoriamente mediante: el parámetro a con distribución $U[0.5, 2]$, el parámetro b con distribución $U[-2.5, 2.5]$ y el parámetro c con distribución $U[0, 0.25]$. Se realizaron 1000 réplicas para cada combinación de valores.

Primeramente veamos el efecto de la cantidad de ítems en la desviación estándar de las estimaciones. Se graficó para un punto fijo la desviación estándar de la estimación en 1000 réplicas para 500, 750 y 1000 sujetos y 10, 25, 50, 75 y 100 ítems. Adicionalmente, en el caso de 1000 sujetos se simuló también 40 y 125 ítems.

A partir de lo anterior se puede pensar una relación del estilo $s = \alpha n^\beta$ con $\beta < 0$ y donde s indica la desviación y n la cantidad de ítems. Si realizamos la transformación logarítmica, obtenemos que la correlación entre ambas variables es -0.97, para el caso de 1000 sujetos. Si analizamos el efecto de la cantidad de sujetos para n fijo, encontramos que la correlación entre N y s es -0.24.

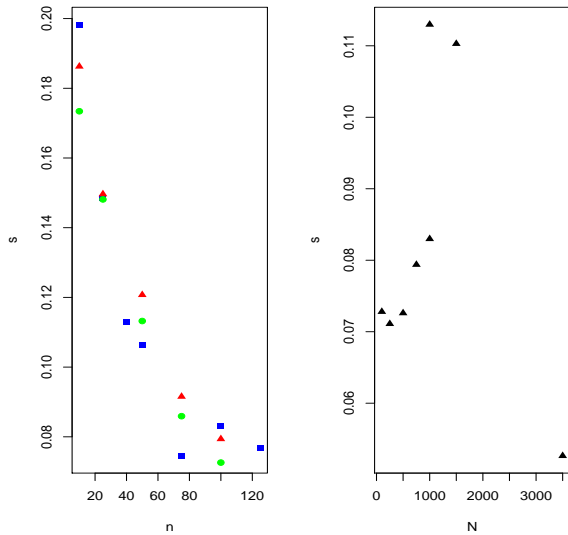


Figura 5.5: DT por ítems y sujetos

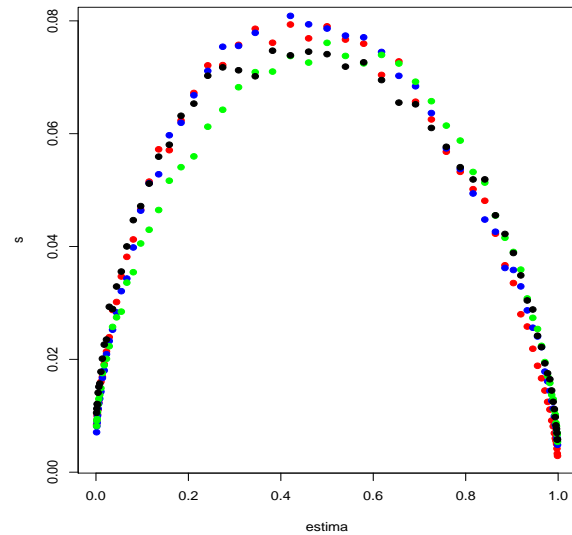


Figura 5.6: DT y estimado

La figura 5.5 muestra primeramente un gráfico con las desviaciones de la estimación del rasgo cuando se consideran 1000 sujetos (azul), 750 sujetos (rojo) y 500 sujetos (verde) para diferentes cantidades de ítems en el test. El otro gráfico contiene las desviaciones para un test con 100 ítems y aplicado a diferentes cantidades de sujetos.

La figura 5.6 muestra la desviación en el rango de puntos estimado para 100 ítems y 500 (verde), 750 (azul) y 1000 (rojo) sujetos y para 75 ítems y 1000 sujetos (negro). Claramente, la forma parabólica nos muestra una relación del estilo $s = \lambda(\hat{\theta}(1 - \hat{\theta}))^\delta$. Tomando logaritmos, la correlación obtenida es 0.98.

Se estimó, por lo tanto, el modelo $\log(s) = \alpha \log(n) + \beta \log(\hat{\theta}(1 - \hat{\theta})) + \log(N)$.

Los resultados nos dan una idea de la relación que existe entre N , n y $\hat{\theta}$. Los coeficientes son -0.1404 para $\log(n)$; 0.5066 para $\log(\hat{\theta}(1 - \hat{\theta}))$ y -0.1748 para $\log(N)$. El R^2 ajustado vale 0.9909.

A partir de los datos simulados, se calcularon los intervalos de confianza para la distribución uniforme y, adicionalmente, para la que suponemos es la verdadera. En la figura siguiente vemos los intervalos de confianza del 95 % donde en el primer gráfico el rasgo tiene distribución uniforme y en el segundo gráfico se tiene el caso en el que suponemos que el rasgo tiene distribución normal.

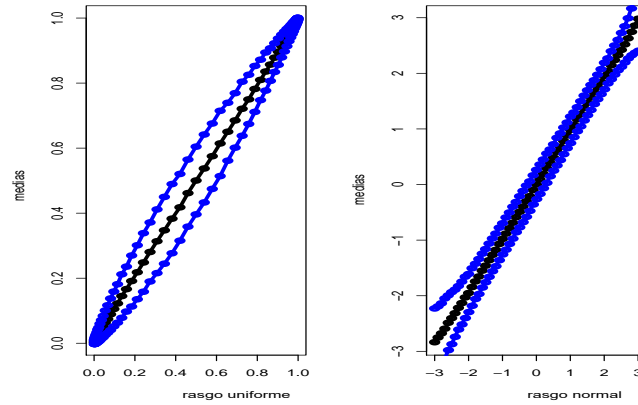


Figura 5.7: Intervalos de confianza

Como hemos probado anteriormente la estimación del rasgo es consistente, lo que se visualiza en los gráficos anteriores al quedar las medias en la diagonal del cuadrado $[0,1] \times [0,1]$. Sin embargo, como hemos afirmado, no hemos podido encontrar la distribución del rasgo estimado. Este estudio nos muestra que la desviación estándar de la estimación no es constante a lo largo del soporte del rasgo. Cuando consideramos el caso canónico (rasgo uniforme) se encuentra que la mayor dispersión se obtiene en los rasgos medios (en el entorno de 0.5), cuando aplicamos la transformación inversa de la normal la menor dispersión se obtiene en el entorno de 0.

Este estudio nos mostró que la desviación depende de la cantidad de ítems, del valor de θ y de la cantidad de sujetos. Es muy interesante que el efecto del valor de θ en la desviación sea aproximadamente $(\theta(1 - \theta))^{1/2}$. El procedimiento que hemos aplicado nos sirve para encontrar empíricamente los intervalos de confianza de los estimadores. Como no conocemos los rasgos verdaderos deberemos utilizar un procedimiento de bootstrap. Finalmente, está abierto el problema de encontrar la distribución asintótica del rasgo, pero ya sospechamos qué tipo fórmula debemos encontrar para la varianza asintótica.

5.4. Estudio 3: Efecto del tamaño muestral en la estimación de las CCI unidimensionales.

Para evaluar el efecto del tamaño muestral se calculó el promedio de las distancias mencionadas en un test de 30 ítems logísticos de tres parámetros. Los parámetros se simulaban como ya hemos indicado en la sección 5.1.4. Se aplicó el test a 100, 250, 500 y 1000 individuos cuyos rasgos provienen de una distribución normal $(0,1)$. Se utilizó un ancho de ventana igual a 0.5 y un núcleo normal.

La figura que sigue muestra las CCI estimadas y real de un ítem. Este ítem tiene los siguientes parámetros ($a=0.65$, $b=-0.18$, $c=0.22$) y su CCI está graficada en color negro. Se muestran también las CCI estimadas para 4 tamaños muestrales 100 (rojo), 250 (verde), 500 (azul) y 1000 (celeste).

El cuadro que sigue muestra los resultados por tamaño muestral en tres medidas del ajuste de las CCI (RECI, RECE y d3) dadas por la definición 5.1.1. Los resultados obtenidos fueron:

N	RECI	RECE	d3
100	0.1736	0.0684	0.1197
250	0.1353	0.0493	0.0753
500	0.1688	0.0552	0.0941
1000	0.1357	0.0480	0.0916

Cuadro 5.4: Ajuste y tamaño muestral

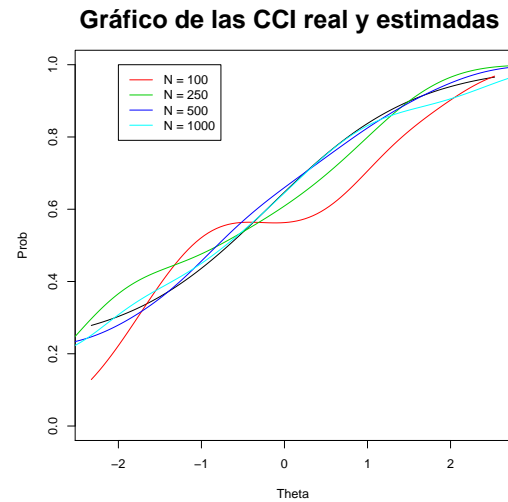


Figura 5.8: Ajuste y tamaño muestral

Los resultados muestran un buen ajuste desde tamaños medianos (250) de la muestra. Esto nos indica que este es un procedimiento que compite con los paramétricos en lo que refiere al tamaño muestral necesario para un buen ajuste.

5.5. Estudio 4: Tamaño muestral en dos dimensiones. Caso I.

En este caso I se analizan ítems dependientes de un solo rasgo, es decir, se consideraron rasgos bivariados pero los ítems dependen de uno de ellos. Se simularon test de 60 ítems, en los que los primeros 30 ítems dependen de un rasgo y los siguientes 30 de otro, siguiendo un modelo 3P cada uno. Los parámetros de los ítems se generaron como está indicado en la sección 5.1.4. Los rasgos se simulaban a partir distribuciones bivariadas con marginales conocidas. En este estudio los rasgos provienen de una cópula con marginales normales(0,1) y correlación de 0.75.

Se utilizaron tamaños muestrales de 100, 500 y 1000 sujetos. Para la estimación no paramétrica se utilizó un núcleo normal y una matriz de ancho de ventana diagonal. Las ventanas de la diagonal se fijaron iguales a 0,2.

Para evaluar el efecto de distintos tamaños de muestra se calculó el promedios de las distancias RECI, RECE y d3.

N	RECI	RECE	d3
100	0.1112	0.0232	0.2565
500	0.0224	0.0042	0.2224
1000	0.0104	0.0021	0.1760

Cuadro 5.5: Ajuste en dos dimensiones. Caso I

Las gráficas que siguen muestran un ejemplo de los ítems simulados. El ítem, en este caso, tiene parámetros: $a=1.65$, $b=-0.75$ y $c=0.23$. Se presenta el modelo real y el estimado usando un tamaño muestral de 500. Como se observa, al depender el ítem de una sola dimensión, las curvas de nivel en la otra dimensión son paralelas. En la superficie estimada vemos que se pueden

producir distorsiones de este paralelismo debido a que el estimador considerado no tiene esa restricción incorporada.

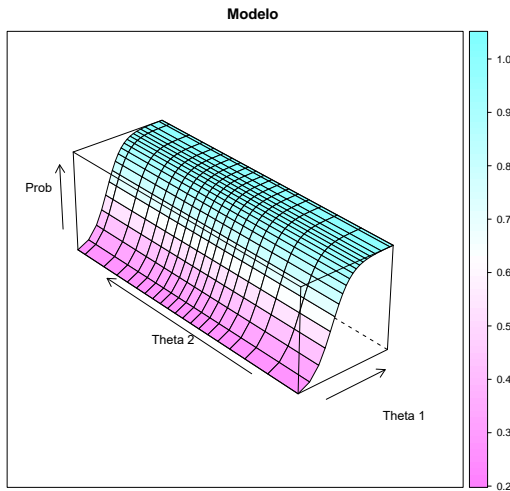


Figura 5.9: Modelo real

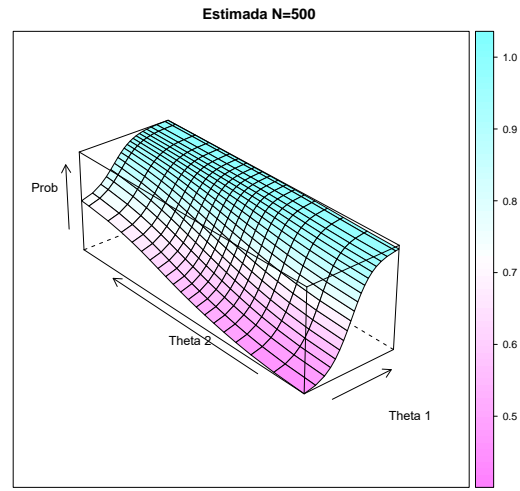


Figura 5.10: Estimado

La tabla nos muestra buenos ajustes a partir de tamaños medianos de muestra (500), considerando que estamos en dos dimensiones. Los resultados en la distancia d_3 también nos muestran un buen ajuste pues el máximo error es pequeño. Sin embargo, se observa que la ganancia en exactitud en esta medida no es tan grande como en las otras medidas. Esto se debe a que, en general, siempre existe algún punto donde el error puede ser grande, pero cuando se comparan las medidas ponderadas para diferentes tamaños muestrales observamos que este tipo de puntos disminuye sustancialmente.

5.6. Estudio 5: Tamaño muestral en dos dimensiones. Caso II.

En este estudio se consideraron ítems dependientes de dos rasgos, como pueden ser idioma español y matemática. Parece lógico suponer que existe correlación entre los niveles de conocimiento de distintos rasgos, por lo que, para simular estos rasgos, utilizaremos distribuciones bivariadas con marginales conocidas y distintos grados de correlación. El modelo que utilizaremos será el logístico multivariado de tres parámetros. Los parámetros de los ítems se generaron como está indicado en la sección 5.1.4 .

Se evaluaron los efectos de distintos tamaños muestrales, para ello se aplicó un test de 60 ítems a 100, 500 y 1000 individuos cuyos rasgos fueron simulados de una distribución bivariada con marginales $N(0,1)$. En este caso los ítems dependen de las dos dimensiones. En la regresión no paramétrica se utilizó un ancho de ventana igual a 0.2 (como en el estudio 4) y un núcleo normal.

Las gráficas que siguen muestran un ejemplo de los ítems simulados. El ítem en este caso tiene parámetros: $a_1=1.28$, $a_2=1.21$, $b=1.06$ y $c=0.13$. Se usó una cópula con marginales normales. Se presenta el modelo real y el estimado usando un tamaño muestral de 500. Los ejes x e y representan los rasgos y el eje z la probabilidad.

Marginales	N	RECI	RECE	D3
$N(0,1);N(0,1)$	100	0.1219	0.0238	0.3000
$N(0,1);N(0,1)$	500	0.0287	0.0053	0.3062
$N(0,1);N(0,1)$	1000	0.0160	0.0029	0.3456

Cuadro 5.6: Ajuste de las superficies

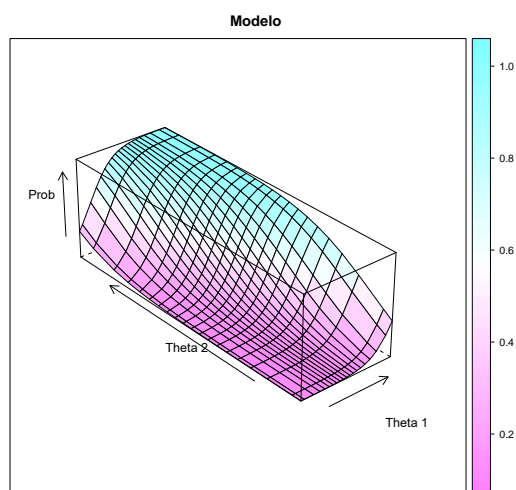


Figura 5.11: Real

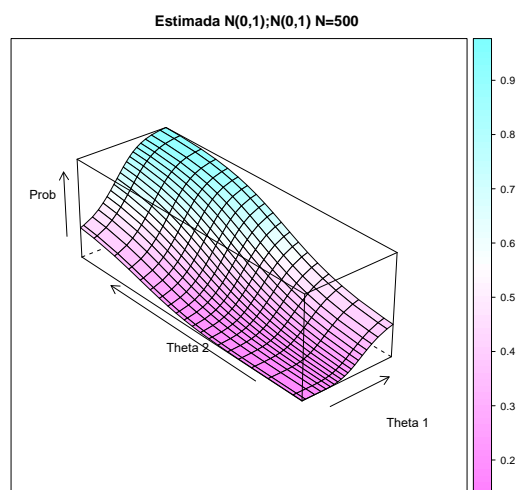


Figura 5.12: Estimado

En este caso también se corroboró un buen ajuste y la necesidad de tamaños muestrales mayores, pero ya con 500 se tiene un buen ajuste. El mismo comentario realizado en la sección anterior vale para la distancia d_3 .

5.7. Estudio 6: Ajuste por ítems y sujetos en dos dimensiones.

A los efectos de ver un poco más la relación del ajuste con el tamaño del test y el tamaño muestral simulamos test bivariados con 10, 20, 50 y 100 ítems aplicados a 100, 250, 500, 750 y 1000 sujetos. Los rasgos se simularon a través de una cópula normal con marginales normales estandar y con los rasgos independientes. Los parámetros se simularon como ya hemos indicado en la sección 5.1.4. Se usó un núcleo normal y la matriz de ancho de ventanas como en los estudios 4 y 5. A partir de esto se calculó el promedio de la RECE del modelo no paramétrico. La tabla que sigue nos muestra los resultados obtenidos.

sujetos	ítems	RECE
100	10	0.17 (0.07)
100	20	0.13 (0.05)
100	50	0.10 (0.04)
100	100	0.12 (0.05)
250	10	0.09 (0.08)
250	20	0.11 (0.06)
250	50	0.09 (0.05)
250	100	0.10 (0.05)
500	10	0.15 (0.08)
500	20	0.11 (0.09)
500	50	0.09 (0.06)
500	100	0.07 (0.05)
750	10	0.21 (0.09)
750	20	0.15 (0.04)
750	50	0.10 (0.05)
750	100	0.08 (0.05)
1000	10	0.23 (0.10)
1000	20	0.16 (0.06)
1000	50	0.11 (0.04)
1000	100	0.07 (0.06)

Cuadro 5.7: Ajuste según muestra y cantidad de ítems

Vemos que tenemos un buen ajuste. Se observa, como es de esperar, la tendencia que si el número de ítems sube, el número de sujetos debe crecer para obtener ajustes adecuados. Debemos recordar que debe mantenerse una relación entre la cantidad de sujetos y de ítems para obtener la consistencia. Específicamente la hipótesis 7 para la consistencia conjunta nos dice que $H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}} < N < Mn^r$ con $r > d + 1$, ver 3.7 .

5.8. Estudio 7: Efecto de la elección de la ventana en una dimensión.

Si bien hemos trabajado en presentar métodos para la elección del mejor ancho de ventana, queremos analizar qué sucede al tomarse diferentes ventanas. Como ya sabemos es crucial la elección de las mismas debido al efecto compensatorio que tienen entre el sesgo y la varianza. Para ello se simuló tests considerando un rasgo normal y se estimaron las CCI con diferentes ventanas.

Se tomaron valores iguales a 0.04, 0.06, 0.08, 0.10, 0.12, 0.14 y 0.16. De cada uno de estos valores de ventana se simuló test de 10, 20, 50, 100 y 150 ítems que eran suministrados a 100, 250, 500, 750, 1000, 2000 y 5000 sujetos en todas las combinaciones posibles, es decir, 35 tipos de simulaciones para cada tipo de ventana. Como dijimos, el rasgo se simuló mediante la distribución normal y la ítems siguieron dos modelos, el modelo 3P y el logit cúbico, donde los parámetros se simuló como está explicado en la sección 5.1.4 .

Se calcularon los valores promedios de RECI, RECE, DAM y D3 para la estimación no paramétrica basada en núcleos ecuación 3.7 y la estimación basada en polinomios locales de grado 1 con núcleos, dada por la ecuación 3.11.

Las tablas que siguen muestran los resultados para las ventanas utilizadas. En las tablas la primera columna indica el ancho de ventana (h), las columnas que siguen tienen la medida de ajuste utilizada (RECI, RECE, DAM y D3) seguida de NP (no paramétrica) para indicar que nos referimos a la estimación mediante la ecuación 3.7 o seguida de POL (no paramétrica por polinomios locales de grado 1) para indicar que nos referimos a la estimación mediante la ecuación 3.11.

h	RECINP	RECIPO	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
0.04	0.30	0.29	0.12	0.12	0.47	0.47	0.37	0.37
0.06	0.29	0.28	0.13	0.13	0.46	0.46	0.33	0.33
0.08	0.26	0.26	0.11	0.12	0.42	0.42	0.28	0.28
0.10	0.24	0.24	0.11	0.11	0.39	0.39	0.25	0.25
0.12	0.23	0.23	0.10	0.10	0.37	0.37	0.22	0.22
0.14	0.22	0.22	0.10	0.10	0.36	0.36	0.20	0.21
0.16	0.21	0.21	0.09	0.09	0.35	0.35	0.19	0.19

Cuadro 5.8: Media de los ajustes según ventana

h	RECINP	RECIPO	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
0.04	0.28	0.28	0.10	0.10	0.45	0.45	0.36	0.36
0.06	0.26	0.26	0.09	0.09	0.44	0.44	0.34	0.33
0.08	0.23	0.23	0.08	0.08	0.39	0.39	0.28	0.27
0.10	0.21	0.21	0.07	0.07	0.35	0.35	0.23	0.23
0.12	0.19	0.19	0.06	0.07	0.33	0.33	0.20	0.20
0.14	0.18	0.18	0.06	0.06	0.30	0.31	0.17	0.18
0.16	0.17	0.17	0.06	0.06	0.29	0.29	0.16	0.16

Cuadro 5.9: Media de los ajustes modelos logísticos según ventana

h	RECINP	RECIPO	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
0.04	0.33	0.33	0.17	0.17	0.52	0.52	0.38	0.38
0.06	0.31	0.31	0.16	0.16	0.48	0.48	0.33	0.33
0.08	0.29	0.29	0.15	0.15	0.45	0.45	0.29	0.29
0.10	0.27	0.27	0.14	0.14	0.43	0.43	0.27	0.27
0.12	0.26	0.26	0.14	0.14	0.42	0.42	0.25	0.25
0.14	0.26	0.26	0.14	0.14	0.41	0.41	0.23	0.23
0.16	0.25	0.25	0.13	0.13	0.40	0.40	0.22	0.22

Cuadro 5.10: Media de los ajustes modelos logit cúbicos según ventana

Se observa que, tanto en el promedio utilizando todos los datos, como cuando se analiza el ajuste en los modelos logísticos y logit cúbicos, hay una relación marcada entre el ajuste y el valor de h .

También se observa que el ajuste mejora a medida que estamos aumentando la ventana por lo que planteamos una segunda serie de simulaciones para encontrar hasta donde mejora el ajuste. Para ello, utilizando un test de 30 ítems de modelo 3P aplicado a 1000 individuos, estudiamos el efecto de considerar distintos anchos de ventana mayores a los anteriores. Usamos ahora 5

ventanas de valores 0.2; 0.4; 0.6; 0.8; 1. Se observa que el mejor ajuste se da en la ventana de 0.4 y luego comienza a crecer.

En la figura 5.13 se muestra un ítem de parámetros $a=0.67$, $b=-0.18$ y $c=0.22$. En negro se muestra la CCI real. También se graficaron las CCI estimadas para las ventanas: $h=0.2$ (rojo), $h=0.4$ (verde), $h=0.6$ (azul), $h=0.8$ (celeste) y $h=1$ (lila).

	RECINP	RECENP	D3NP
$h = 0.2$	0.1302	0.0339	0.1005
$h = 0.4$	0.1119	0.0368	0.0784
$h = 0.6$	0.1773	0.0622	0.1115
$h = 0.8$	0.2854	0.0916	0.1611
$h = 1$	0.3935	0.1174	0.2092

Cuadro 5.11: Ajuste medio por ventana

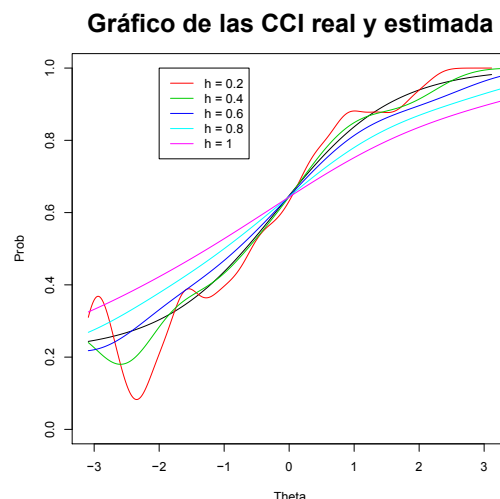


Figura 5.13: CCI para diferentes ventanas

Esta serie de simulaciones nos muestra la importancia de la elección de la ventana. Ya hemos mencionado que existe un compromiso permanente entre sesgo y varianza. El sesgo disminuye si la ventana es pequeña y en cambio, la varianza decrece si la ventana es grande.

Como vemos, esto se ve reflejado en los datos del ajuste que hemos calculado para las diferentes ventanas.

Nosotros en la sección 3.10.1 mostramos el método de validación cruzada para la elección de la ventana. Hay otros métodos que pueden verse en Fan y Gijbels (1996). Por otro lado, Lee (2007) utiliza un método más cualitativo al elegir como ventana aquella que hace la curva más suave.

5.9. Estudio 8: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso I.

Para comparar las estimaciones mediante regresión no paramétrica al variar los anchos de ventana consideramos un test de 60 ítems aplicado a 500 individuos cuyos rasgos fueron simulados mediante una distribución bivariada con marginales $N(0,1)$ independientes, usando una cópula normal. Los 60 ítems se dividieron en dos grupos de 30 donde cada grupo depende de uno solo de los rasgos. Los parámetros de los ítems fueron generados como está explicado en la sección 5.1.4 .

Luego, realizamos las estimaciones utilizando distintos anchos de ventana (la matriz H se considera diagonal) y calculamos las tres distancias ya definidas. En todos los casos, utilizamos un núcleo normal en la regresión no paramétrica. La tabla muestra los resultados obtenidos, significando los encabezados lo mismo que en la sección anterior. El mejor ancho de ventana para las diagonales de la matriz H es 0.2, aunque no hemos probado matrices con valores de la

ventana menores.

h	RECINP	RECENP	D3NP
0.2	0.0244	0.0044	0.2490
0.4	0.0303	0.0059	0.2415
0.6	0.0351	0.0069	0.2832
0.8	0.0353	0.0073	0.2278
1	0.0428	0.0083	0.2627

Cuadro 5.12: Ajuste por matriz de ventanas. Caso I

5.10. Estudio 9: Efecto de distintos valores de ancho de ventana en dos dimensiones. Caso II.

Utilizamos el mismo tipo de test que en la sección anterior, o sea, un test de 60 ítems aplicado a 500 individuos cuyos rasgos fueron simulados mediante una distribución bivariada con marginales $N(0,1)$ independientes. Ahora los 60 ítems dependen de ambos rasgos y sus parámetros se simularon como está explicado en la sección 5.1.4 .

Calcularemos las tres distancias considerando distintos anchos de ventana. Continuamos utilizando en la regresión no paramétrica un núcleo normal. Los resultados son iguales a las del estudio 8 y realizamos la misma observación respecto a la ventana que tiene mejor ajuste.

h	RECINP	RECENP	D3NP
0.2	0.0288	0.0054	0.2918
0.4	0.0444	0.0100	0.3876
0.6	0.0552	0.0135	0.3870
0.8	0.0636	0.0150	0.4507
1	0.0663	0.0160	0.4454

Cuadro 5.13: Ajuste por matriz de ventanas. Caso II

Finalmente, podemos evidenciar a partir de los estudios 7, 8 y 9 como el ajuste se ve influenciado por el ancho de ventana. Es por este motivo que desarrollar métodos simples para la elección de la ventana óptima es muy importante. Como hemos mencionado antes, en la tesis se presentan métodos para la elección de la ventana, pero están referidos a cuando se considera el rasgo estimado sin error. Si queremos considerar el error en la estimación del rasgo, deberemos generar un método que lo contemple. Retomaremos esto en el capítulo de conclusiones, pues este aspecto es claramente una limitación y un desafío para un desarrollo futuro.

5.11. Estudio 10: Efecto del tipo de núcleo en una dimensión.

En esta sección presentamos los resultados de la utilización de diferentes núcleos en la estimación de las CCI. Pretendemos mostrar que los estimadores funcionan bien con cualquier tipo de núcleo, aun con los que no tienen soporte compacto como ser el núcleo gaussiano.

Se realizaron estimaciones con 10 núcleos distintos usando diferentes modelos, tamaños muestrales y cantidad de ítems. Se utilizaron 4 modelos de TRI (1P,2P,3P, L3) descritos en la sección 5.1.3. Para cada núcleo se realizaron 140 simulaciones dadas de combinar los 4 modelos, tests de 10, 20, 50, 100, 150 ítems y tamaños muestrales de 100, 250, 500, 750, 1000, 2000 y 5000 sujetos. Los rasgos se simulaban normales estándar y la ventana se fijó en 0.2 debido a su buen funcionamiento en los estudios previos. Se usaron como estimadores de las CCI los propuestos por las ecuaciones 3.7 (NP) y 3.11 (POL).

Las tres tablas que siguen muestran las medias de las medidas de ajuste de los diferentes núcleos. La primera tabla resume las medias de los ajustes por núcleo para todos los modelos, la segunda considera solo los modelos logísticos, y la tercera, los modelos logit cúbicos. Los encabezados de las tablas tienen la misma codificación que en las secciones anteriores.

núcleo	RECINP	RECIPOL	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
normal	0.14	0.15	0.05	0.06	0.23	0.24	0.12	0.12
epanechnikov	0.23	0.23	0.10	0.10	0.38	0.38	0.24	0.24
uniforme	0.22	0.22	0.09	0.09	0.36	0.36	0.23	0.24
triangular	0.24	0.24	0.11	0.10	0.39	0.39	0.25	0.25
biweight	0.24	0.24	0.11	0.10	0.39	0.40	0.26	0.26
triweight	0.25	0.25	0.11	0.11	0.42	0.41	0.28	0.28
tricubo	0.24	0.24	0.11	0.11	0.39	0.39	0.26	0.26
coseno	0.23	0.23	0.10	0.10	0.38	0.38	0.24	0.25
logístico	0.18	0.17	0.08	0.08	0.32	0.30	0.13	0.13
sigmoideo	0.18	0.18	0.08	0.08	0.32	0.30	0.14	0.14

Cuadro 5.14: Media de los ajustes según núcleo

núcleo	RECINP	RECIPOL	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
normal	0.13	0.13	0.04	0.04	0.21	0.21	0.10	0.11
epanechnikov	0.19	0.20	0.06	0.06	0.33	0.33	0.21	0.22
uniforme	0.18	0.18	0.06	0.06	0.30	0.31	0.21	0.22
triangular	0.20	0.20	0.07	0.06	0.35	0.35	0.23	0.23
biweight	0.21	0.21	0.07	0.07	0.36	0.36	0.24	0.25
triweight	0.22	0.22	0.07	0.07	0.38	0.38	0.27	0.27
tricubo	0.21	0.21	0.07	0.07	0.36	0.35	0.24	0.248
coseno	0.2	0.20	0.06	0.06	0.33	0.33	0.22	0.224
logístico	0.12	0.12	0.04	0.04	0.21	0.21	0.08	0.08
sigmoideo	0.12	0.13	0.04	0.04	0.21	0.21	0.10	0.09

Cuadro 5.15: Media de los ajustes modelos logísticos según núcleo

núcleo	RECINP	RECIPOL	RECENP	RECEPOL	DAMNP	DAMPOL	D3NP	D3POL
normal	0.23	0.23	0.12	0.12	0.39	0.38	0.19	0.19
epanechnikov	0.26	0.27	0.14	0.14	0.42	0.42	0.26	0.26
uniforme	0.26	0.26	0.13	0.13	0.41	0.41	0.26	0.26
triangular	0.27	0.27	0.14	0.14	0.43	0.43	0.27	0.27
biweight	0.27	0.27	0.14	0.14	0.44	0.44	0.28	0.28
triweight	0.28	0.28	0.14	0.14	0.45	0.45	0.29	0.29
tricubo	0.27	0.28	0.14	0.14	0.44	0.44	0.28	0.28
coseno	0.27	0.27	0.14	0.14	0.43	0.43	0.27	0.27
logístico	0.24	0.23	0.12	0.12	0.44	0.40	0.18	0.19
sigmoideo	0.24	0.23	0.12	0.12	0.42	0.39	0.19	0.19

Cuadro 5.16: Media de los ajustes modelos logit cúbicos según núcleo

Cuando analizamos las medidas de ajuste, observamos que el ajuste es bueno para todos los núcleos, inclusive el núcleo normal que, aunque no tiene soporte compacto funciona muy bien. Recordemos que la prueba de consistencia conjunta pedía que el núcleo tuviera soporte compacto. En el capítulo de conclusiones retomaremos este aspecto y propondremos una posible hipótesis sobre el núcleo que permite considerar a los de soporte no acotado, dejando la prueba para un desarrollo futuro.

Adicionalmente, notamos que, si bien se tiene un mayor error (aunque pequeño) en los modelos logit cúbicos es igual para los diferentes núcleos.

Podemos, entonces, concluir que el método es robusto ante el uso de diferentes núcleos unidimensionales.

5.12. Estudio 11: Efecto de distintos núcleos en dos dimensiones. Caso I.

Con el fin de evaluar los efectos de distintos núcleos en la regresión no paramétrica consideramos un test de 60 ítems aplicado a 500 individuos cuyos rasgos fueron simulados mediante una distribución bivariada con marginales $N(0,1)$ independientes, usando una cópula normal. Los 60 ítems se dividieron en dos grupos de 30 donde cada grupo depende de uno solo de los rasgos. Los parámetros de los ítems fueron generados como está explicado en la sección 5.1.4.

De los posibles núcleos multidimensionales se utilizaron los siguientes cuatro núcleos multivariados que son clásicos: Normal, Epanechnikov esférico y dos núcleos de Bartlett-Epanechnikov. En todos los casos se fijó la matriz de ancho de ventana con diagonal igual a 0.2 (debido a los resultados del estudio 8 y 9).

1. $(2\pi)^{-d/2} |\mathbf{I}|^{-1/2} e^{-\frac{1}{2} \mathbf{x}^T \mathbf{I}^{-1} \mathbf{x}}$ (normal) con \mathbf{I} la identidad.
2. $(1 - x^t x) \quad |x^t x| \leq 1$ (Epanechnikov esférico)
3. $3\pi^{-1} (1 - x^t x)^2 \quad |x^t x| \leq 1$ (K_2)
4. $4\pi^{-1} (1 - x^t x)^3 \quad |x^t x| \leq 1$ (K_3)

Los resultados fueron:

	RECINP	RECENP	D3NP
normal	0.0104	0.0021	0.1760
Epanechnikov	0.0092	0.0019	0.2184
K_2	0.0095	0.0018	0.2572
K_3	0.0109	0.0021	0.3218

Cuadro 5.17: Ajuste por núcleo caso I

Nuevamente observamos el buen ajuste que tienen los diferentes núcleos.

5.13. Estudio 12: Efecto de distintos núcleos en dos dimensiones. Caso II.

Con el fin de evaluar los efectos de distintos núcleos en la regresión no paramétrica consideramos el mismo tipo de test del estudio anterior, o sea, un test de 60 ítems aplicado a 500 individuos cuyos rasgos fueron simulados mediante una distribución bivariada con marginales $N(0,1)$ independientes. La diferencia es que, en este caso, los 60 ítems dependen de ambos rasgos y sus parámetros se simularon como está explicado en la sección 5.1.4.

Se utilizaron los mismos núcleos que en el estudio 11, e igualmente en todos los casos fijamos el ancho de ventana igual a 0.2, por lo comentado anteriormente.

	media de d1	media de d2	media de d3
normal	0.0287	0.0053	0.3062
Epanechnikov	0.0895	0.0103	0.6855
K_2	0.0895	0.0101	0.6867
K_3	0.0919	0.0110	0.6985

Cuadro 5.18: Ajuste por núcleo caso II

A partir de los estudios 10,11 y 12 podemos corroborar la robustez de los estimadores ante los diferentes núcleos, tanto en el caso unidimensional, como en el bidimensional. Además el estimador funciona muy bien con núcleos que no tienen soporte compacto. Esto es muy bueno, ya que el poder usar, por ejemplo, el núcleo normal nos permite obtener estimaciones suaves de las curvas.

5.14. Estudio 13: Efectos de la distribución de los rasgos en una dimensión

Como hemos demostrado, las CCI permanecen incambiadas ante transformaciones monótonas del rasgo, por lo tanto, la distribución del rasgo no debería afectar la estimación de los modelos. Para corroborar este hecho se realizó un estudio del ajuste ante las diferentes distribuciones. Se consideraron las distribuciones: normal, beta, arcoseno, logística, lognormal, normal sesgada y uniforme.

Se simularon, de cada una de estas distribuciones, 500 sujetos y se les aplicó test de 10, 20, 50, 100 y 150 ítems que siguen un modelo 3P con parámetros simulados, como está explicado en la sección 5.1.4.

Para estimar las CCI se usó un núcleo normal y la ventana como lo venimos haciendo se tomó de 0.20.

Las tablas que siguen nos muestran las medidas de ajuste ante las diferentes distribuciones del rasgo.

Distribución	RECI NP	RECI POL	RECE NP	RECE POL	D3 NP	D3 POL
normal	0.24 (0.08)	0.25 (0.08)	0.05 (0.01)	0.06 (0.01)	0.12 (0.09)	0.13 (0.08)
beta	0.23(0.08)	0.26 (0.10)	0.08 (0.01)	0.08 (0.01)	0.13(0.07)	0.15(0.09)
arcoseno	0.28 (0.11)	0.27 (0.10)	0.06 (0.03)	0.08 (0.02)	0.21 (0.08)	0.23 (0.10)
logística	0.29 (0.12)	0.3 (0.14)	0.08 (0.01)	0.08 (0.02)	0.20 (0.08)	0.20 (0.08)
lognormal	0.27 (0.11)	0.31 (0.14)	0.08 (0.02)	0.08 (0.02)	0.19 (0.09)	0.21 (0.09)
normal sesgada	0.25 (0.09)	0.24 (0.09)	0.05 (0.01)	0.08 (0.01)	0.15 (0.09)	0.18 (0.10)
uniforme	0.24 (0.08)	0.24 (0.08)	0.06 (0.01)	0.06 (0.01)	0.11 (0.08)	0.11 (0.08)

Cuadro 5.19: Ajuste según distribución

La tabla nos corrobora que, los valores de ajuste de los estimados de las CCI cuando el rasgo sigue diferentes distribuciones, es aproximadamente igual en todas ellas.

5.15. Estudio 14: Efecto de distintas distribuciones marginales de θ en dos dimensiones. Caso I.

Se evaluó el efecto de considerar distintas distribuciones bivariadas para los rasgos de los individuos. Para ello utilizamos diferentes distribuciones marginales y fijamos el nivel de correlación entre ambas igual a 0,75. Las cópulas consideradas tienen las siguientes marginales: $N(0,1)$ y $N(0,1)$; $Beta(5,5)$ y $Beta(5,5)$; $U(0,1)$ y $U(0,1)$. Se transforma cada marginal a los efectos de hacer comparables los resultados.

Se simula un test de 60 ítems, que se dividieron en dos grupos de 30, donde cada grupo depende de uno solo de los rasgos. Los parámetros de los ítems fueron generados como está explicado en la sección 5.1.4. Para estimar las CCI se utilizó un núcleo normal y una matriz diagonal de ventanas con valores iguales a 0.20.

Obtenemos:

	RECINP	RECENP	D3NP
$N(0,1);N(0,1)$	0.0244	0.0040	0.1874
$Beta(5,5);Beta(5,5)$	0.0264	0.0047	0.2526
$U(0.1,0.9);U(0.1,0.9)$	0.0155	0.0028	0.2373

Cuadro 5.20: Ajuste por distribución marginal caso I

Se observa que el método funciona muy bien con cualquiera de los pares de marginales que se trabajó.

5.15.1. Distintos niveles de correlación entre los rasgos

Ahora consideramos distintos niveles de correlación entre las distribuciones marginales ($N(0,1)$) que componen la cópula, es decir, consideraremos distintos niveles de correlación entre las dos rasgos de cada individuo. Se tomarán correlaciones 0 (dos rasgos independientes), 0.2375, 0.475, 0.7125, 0.95 (casi un único rasgo). El caso en el cual el coeficiente de correlación es igual a cero implica que se está suponiendo independencia entre los dos rasgos en cuestión. A continuación, mostramos la tabla con los resultados.

rho	RECINP	RECENP	D3NP
0	0.0092	0.0017	0.1150
0.2375	0.0091	0.0017	0.1291
0.475	0.0114	0.0018	0.2259
0.7125	0.0132	0.0021	0.2262
0.95	0.0142	0.0026	0.2751

Cuadro 5.21: Ajuste por correlación entre rasgos caso I

También queda claro que el ajuste es bueno ante diferentes correlaciones entre los rasgos.

5.16. Estudio 15: Efecto de distintas distribuciones de θ en dos dimensiones. Caso II.

Se realiza un estudio análogo al estudio 14, con la diferencia que ahora los 60 ítems dependen de los dos rasgos. Se usa un modelo logístico multidimensional con los parámetros simulados como en la sección 5.1.4.

	RECINP	RECENP	D3NP
N(0,1);N(0,1)	0.0154	0.0029	0.2457
Beta(5,5);Beta(5,5)	0.0308	0.0060	0.3128
U(0,1);U(0,1)	0.0179	0.0025	0.2858

Cuadro 5.22: Ajuste por distribución marginal caso II

También en este caso, el ajuste es muy bueno ante las diferentes combinaciones de marginales.

5.16.1. Efecto de distintos grados de correlación entre los rasgos

Para evaluar el efecto de distintos grados de correlación entre los distintos rasgos consideramos una distribución bivariada con marginales con distribución N(0,1) y calculamos las tres distancias para distintos niveles de correlación. La tabla que sigue muestra los resultados que son muy buenos.

rho	RECINP	RECENP	D3NP
0	0.0163	0.0038	0.2330
0.2375	0.0153	0.0033	0.2423
0.475	0.0156	0.0032	0.2859
0.7125	0.0159	0.0029	0.3288
0.95	0.0138	0.0022	0.2603

Cuadro 5.23: Ajuste por correlación entre rasgos. Caso II

A partir de los estudios 13,14 y 15 podemos evidenciar que el método propuesto es robusto ante las diferentes distribuciones del rasgo en el caso unidimensional. También es robusto ante diferentes marginales de los rasgos multivariados, así como ante diferentes correlaciones entre los rasgos.

5.17. Estudio 16: Efecto de la elección de la ventana en el modelo isótono

Para este estudio se utilizaron valores de h iguales a: 0.04, 0.06, 0.08, 0.10, 0.12, 0.14 y 0.16. De cada uno de estos valores de ventana se simuló test de 10, 20, 50, 100 y 150 ítems que eran suministrados a 100, 250, 500, 750, 1000, 2000 y 5000 sujetos en todas las combinaciones posibles, es decir, 35 tipos de simulaciones para cada tipo de ventana. El rasgo se simuló mediante la distribución normal y los ítems siguieron dos modelos, el modelo 3P y el logit cúbico, donde los parámetros se simuló como está explicado en la sección 5.1.4.

Se calcularon los valores promedios de RECI, RECE y DAM para la estimación no paramétrica isótoma.

h	RECI	RECE	DAM
0.04	0.50	0.24	0.76
0.06	0.50	0.25	0.80
0.08	0.49	0.24	0.79
0.10	0.49	0.24	0.79
0.12	0.49	0.24	0.79
0.14	0.49	0.24	0.80
0.16	0.49	0.24	0.81

Cuadro 5.24: Media de los ajustes según ventana

h	RECI	RECE	DAM
0.04	0.49	0.23	0.74
0.06	0.50	0.24	0.79
0.08	0.49	0.23	0.78
0.10	0.49	0.23	0.79
0.12	0.49	0.23	0.79
0.14	0.49	0.24	0.80
0.16	0.49	0.24	0.80

Cuadro 5.25: Media de los ajustes modelos logísticos según ventana

h	RECI	RECE	DAM
0.04	0.51	0.26	0.83
0.06	0.50	0.25	0.82
0.08	0.49	0.25	0.80
0.10	0.49	0.25	0.80
0.12	0.49	0.25	0.80
0.14	0.49	0.25	0.80
0.16	0.49	0.25	0.81

Cuadro 5.26: Media de los ajustes modelos logit cúbicos según ventana

En la estimación de la densidad se observó que los ajustes son aproximadamente de la misma magnitud ante diferentes elecciones de la ventana. Este resultado es coherente, ya que el tamaño de la ventana depende del núcleo, de la cantidad de puntos presentes y de la densidad de la variable. El modelo isótono se estima en $[0, 1]$ y se supone una distribución uniforme del rasgo. Como la grilla se establece manualmente, y se hace lo suficientemente densa, no existirá mucha variación en la estimación ante cambios de ventana, excepto en los extremos.

5.18. Estudio 17. Efecto de la elección del núcleo en el modelo isótono

En este estudio pretendemos mostrar que el estimador isótono funcionan bien con cualquier tipo de núcleo, aun con los que no tienen soporte compacto, como ser el núcleo gaussiano.

Se realizaron estimaciones con 10 núcleos distintos usando diferentes modelos, tamaños muestrales y cantidad de ítems. Se utilizaron 4 modelos de TRI (1P,2P,3P, L3) descritos en la sección 5.1.3. Para cada núcleo se realizaron 140 simulaciones obtenidas al combinar los 4 modelos, tests de 10, 20, 50, 100, 150 ítems y tamaños muestrales de 100, 250, 500, 750, 1000, 2000 y 5000 sujetos. Los rasgos se simulaban normales estándar y la ventana se fijó en 0.10 debido a su buen funcionamiento en el caso isótono.

Las tres tablas que siguen muestran las medias de las medidas de ajuste de los diferentes núcleos. La primera tabla resume las medias de los ajustes por núcleo para todos los modelos, la segunda considera solo los modelos logísticos, y la tercera, los modelos logit cúbicos. Los encabezados de las tablas tienen la misma codificación que en las secciones anteriores.

núcleo	RECI	RECE	DAM
normal	0.43	0.20	0.67
epanechnikov	0.52	0.27	0.82
uniforme	0.51	0.26	0.81
triangular	0.57	0.30	0.89
biweight	0.52	0.27	0.82
triweight	0.53	0.27	0.83
tricubo	0.52	0.27	0.82
coseno	0.52	0.27	0.82
logistico	0.43	0.21	0.71
sigmoideo	0.43	0.21	0.70

Cuadro 5.27: Media de los ajustes según núcleo

núcleo	RECI	RECE	DAM
normal	0.43	0.20	0.67
epanechnikov	0.50	0.25	0.77
uniforme	0.49	0.25	0.76
triangular	0.537	0.281	0.84
biweight	0.50	0.26	0.78
triweight	0.50	0.26	0.78
tricubo	0.51	0.26	0.78
coseno	0.502	0.26	0.77
logistico	0.42	0.20	0.69
sigmoideo	0.41	0.20	0.68

Cuadro 5.28: Media de los ajustes modelos logísticos según núcleo

núcleo	RECI	RECE	DAM
normal	0.43	0.22	0.71
epanechnikov	0.53	0.28	0.86
uniforme	0.53	0.27	0.85
triangular	0.60	0.32	0.95
biweight	0.54	0.28	0.86
triweight	0.54	0.28	0.87
tricubo	0.54	0.28	0.86
coseno	0.54	0.28	0.86
logístico	0.44	0.22	0.72
sigmoideo	0.44	0.22	0.72

Cuadro 5.29: Media de los ajustes modelos logit cúbicos según núcleo

Como ocurrió cuando se estudiaron los estimadores no monótonos en el caso del modelo isótono, al utilizar diferentes núcleos, se encontró que el modelo ajustó adecuadamente. Claramente al estar el modelo restringido a ser monótono hay mayor error que cuando se consideran los estimadores no paramétricos y por polinomios locales. Esto se ve reflejado en las medidas de ajuste.

Capítulo 6

Aplicación a datos reales

6.1. Objetivos

En este capítulo mostraremos la aplicación práctica de los métodos presentados y su utilidad para analizar test de diferentes características. Para ello, nos proponemos considerar diferentes condiciones como ser la dimensionalidad, el tipo de respuesta y la monotonía.

Los objetivos específicos son:

- 1- Estimar los rasgos de los sujetos en un test unidimensional.
- 2- Aplicar los cuatro modelos propuestos para estimar las CCI en un test unidimensional.
- 3- Utilizar la regresión no paramétrica para analizar el ajuste de las CCI estimadas paramétricamente.
- 4- Estimar el vector de rasgos de los sujetos en un test con dos dimensiones.
- 5- Estimar las superficies de respuesta al ítem en un test bidimensional.
- 6- Estimar las funciones de respuesta para ítems de respuesta graduada.
- 7- Estimar las superficies de respuesta para cada opción en un test de dimensión media (en nuestro caso 4 dimensiones).
- 8- Representar gráficamente las superficies de respuesta de dimensión mayor a dos, mediante proyecciones y curvas de nivel.

6.2. Métodos

Para cumplir con los objetivos se utilizaron dos test: la evaluación diagnóstica en Matemática al ingreso a la Universidad de la República y el Children Depression Inventory (CDI). En los objetivos 1 a 5 se utilizó la prueba de evaluación diagnóstica en Matemática y para los objetivos 6 a 8 el CDI.

En el caso de la evaluación diagnóstica en Matemática comenzaremos estudiando la dimensionalidad de la prueba. Para esto se utilizaron varios índices:

- 1- Porcentaje de varianza explicado por cada componente.
- 2- El cociente entre autovalores.
- 3- El screeplot de Catrell.
- 4- Minumum Partial Average (MAP).
- 5- Análisis Paralelo.
- 6- Very simple structure (VSS).

Cuando la prueba se consideró unidimensional se realizaron dos estudios:

a) Se calibró mediante el modelo 2P.

b) Se estimó la densidad del rasgo y las CCI mediante regresión no paramétrica, polinomios locales y los modelos isótonos respectivos.

En el análisis paramétrico se utilizó para medir el ajuste de las personas el estadístico L_0 de Levine y Rubin y el estadístico L_z de Drasgow. Para medir el ajuste de los ítems se calculó el Q_1 de Yen. También se utilizó para el ajuste estimar las CCI mediante regresión no paramétrica y calcular las medidas, RECI, RECE y d_3 .

Cuando se consideró la prueba bidimensional, con los rasgos álgebra y cálculo se estimó la densidad no paramétricamente y las curvas de respuesta mediante los cuatro modelos propuestos en la tesis.

En el caso del CDI se tomó en cuenta la estructura factorial de 4 dimensiones encontradas por Kovacs (1992) y Del Barrio et al (1999). Se estimaron las FRI en cada dimensión mediante dos modelos de respuesta graduada no paramétricos, uno basado en regresión no paramétrica y el otro usando el modelo isótono.

También se estudió el CDI con los modelos no paramétricos en dimensión 4, modelizándose las distintas opciones de respuesta. Luego, se proyectaron las superficies en espacios bidimensionales para su representación gráfica. También se utilizó una representación de las curvas de nivel.

6.3. Evaluación diagnóstica en matemática

En el marco del proceso de descentralización de la Universidad de la República de Uruguay (Udelar) que llevó a la creación de 7 Centros Universitarios ubicados fuera de Montevideo, en el Interior del país, la Comisión Coordinadora del Interior (CCI) creyó oportuno evaluar las competencias de los estudiantes al ingreso a la Universidad en el área de matemática con el fin de adaptar los programas educativos y crear dispositivos de apoyos específicos. Por estos motivos se aplican pruebas diagnósticas basadas en estándares desarrolladas por Rodríguez (2016). La prueba fue aplicada a los 1380 estudiantes que ingresaron a la Universidad de la República en los Centros Universitarios Regionales, distribuidos de la forma en que se observa en el mapa (figura 6.1).

En primer lugar, un grupo de docentes estableció estándares de contenido agrupados en cuatro categorías: números y operaciones, álgebra y funciones, geometría y análisis de datos, estadística y probabilidad. Con la finalidad de contar con ítems ya testeados, se tomó como base la dimensión Matemática del Scholastic Assessment Test (SAT) elaborada por The College Board. Un grupo de docentes preseleccionó 186 ítems, repartidos equilibradamente entre las diferentes categorías. Otro grupo de docentes con experiencia en evaluación de ítems los valoraron y establecieron la proporción de ítems para cada categoría teniendo en cuenta la preponderancia que se les da en los programas curriculares a esos contenidos. La prueba de Matemáticas quedó finalmente conformada por dos cuadernillos con 44 ítems cada uno.

A las escalas se les realizó un serie de análisis psicométricos a efectos de estudiar sus propiedades. En primer lugar, se calcularon los índices de dificultad e índices de discriminación de cada ítem así como la distribución de las respuestas.

La dificultad media de los ítems del cuadernillo 1 de matemática fue de 0.36 (0.02), del cuadernillo 2 fue 0.37 (0.02). Se recomienda que el índice de dificultad sea mayor a 0,5 y que la dificultad media sea 0,62 para ítems de 4 opciones (Abad et al., 2011). Si bien los ítems de estas pruebas tienen 5 distractores, se encuentran alejados de lo recomendado para el índice de dificultad. Fueron calculados los índices de discriminación, en este caso las correlaciones

biserialles puntuales (bivariadas) de los ítems por separado y de los grupos. La discriminación media de los ítems del cuadernillo 1 y 2 de Matemática fue de 0.40 (0.01 y 0.02 respectivamente). La bibliografía sugiere que los ítems que obtengan un índice de discriminación menor a 0,20 sean descartados (Abad et al., 2011 y Schmeiser y Welch, 2006). Este criterio se usó para la eliminación de los ítems que no ajustaban. En esta primera instancia no se eliminaron ítems del cuadernillo 1 de matemática; se sacaron los ítems 9, 11, 19, 28 y 31 del cuadernillo 2 de Matemática. Una vez eliminados estos ítems se calculó el coeficiente alfa de Cronbach para determinar la fiabilidad de las pruebas (Rodríguez, 2017). El alfa de Cronbach de las pruebas de Matemática fueron 0.88 para el cuadernillo 1 y 0.87 para el cuadernillo 2 (Rodríguez, 2017).



Figura 6.1: Muestreo de la prueba

6.3.1. Evidencias sobre la estructura interna

El valor global del índice de adecuación muestral de KMO para el cuadernillo 1 de Matemática es de 0.913 y de 0.899 para el cuadernillo 2, que según la clasificación de Kaiser se puede considerar como “maravillosa”. En el test de esfericidad de Barlett el valor del chi-cuadrado para el cuadernillo 1 de Matemática es 5524.307(946) p-value=0 y para el cuadernillo 2 de Matemática es 4643.2(946) p-value=0. Por lo tanto, no podemos mantener la hipótesis nula de que las variables utilizadas en el análisis no se correlacionarían en la población en la que se ha extraído la muestra. A partir de los dos estudios anteriores se concluye que los datos son adecuados para realizar un análisis factorial. Se realiza un análisis de componentes principales con rotación varimax usando correlaciones tetracóricas. Si bien el método de componentes principales es susceptible de críticas se lo utilizó ya que simplemente se quiere evaluar la dimensionalidad. Por ese motivo también se usó la rotación varimax.

Para evaluar la dimensionalidad se utilizaron varios índices. Para la unidimensionalidad se observó el criterio propuesto por Reckase(1979) donde se pide que el primer componente explique, por lo menos, el 20 por ciento de la varianza. También se observó el cociente entre el primer autovalor y el segundo. Este cociente será indicativo de unidimensionalidad si es

aproximadamente 4 o mayor. Se analizó el método gráfico propuesto por Cattell (1966) mediante el scree test y se tomó en cuenta el índice MAP (Minimum Average Partial) de Velicer (1976) que implica calcular el promedio de las correlaciones parciales al cuadrado después de que cada uno de los k componentes ha sido parcializado. Se observa donde alcanza el mínimo el promedio de las correlaciones parciales, el cual se alcanza cuando la matriz de residuos se aproxima a la identidad (Rodríguez, 2016). Adicionalmente se realizó el análisis paralelo de Horn (1965) y el VSS (very simple structure) de Revelle and Rocklin (1979).

Prueba 1 de matemática

El cuadernillo 1 consta de 44 ítems. El gráfico scree para el cuadernillo 1 muestra que se deberían tomar dos componentes, aunque el primer valor propio es mucho mayor que el segundo. El porcentaje de varianza explicada por el primer factor es 26.37 y el ratio entre los dos primeros valores propios es 3.85. Al realizar el análisis paralelo se obtuvo que la cantidad de componentes es de 2. Cuando se analiza el VSS con complejidad 1 se obtiene un máximo de 0.791 con 1 componente. El criterio MAP da muy semejante con 1, 2 y 3 componentes. Las gráficas se observan en la siguiente figura. Es claro que tenemos resultados que indicarían que lo mejor es tomar dos componentes. Sin embargo hay resultados con los cuales se puede considerar la prueba unidimensional. Como se trata de una prueba para evaluar competencias suele tomarse unidimensionalmente. A los efectos de hacerla efectivamente unidimensional, podríamos aplicar alguna técnica, como por ejemplo, eliminar ítems. Si embargo, ese no es nuestro objetivo, ya que lo que queremos es aplicar nuestros modelos no paramétricos. Entonces, a partir de lo anterior realizaremos dos análisis: uno que considera la prueba es unidimensional y la otra bidimensional.

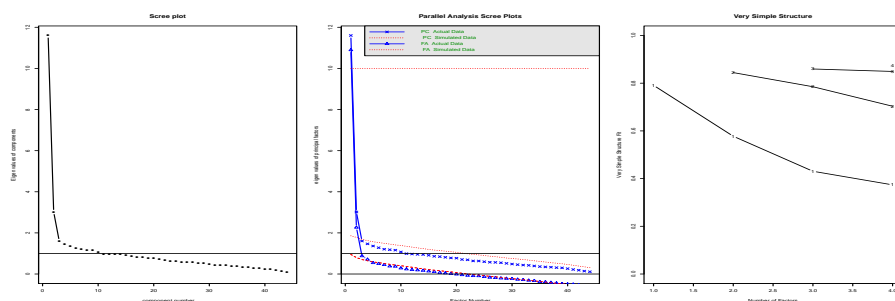


Figura 6.2: Prueba 1-Cantidad de factores

Prueba 2 de matemática

El cuadernillo 2 consta de 39 ítems. El gráfico scree muestra que se deberían tomar dos componentes, el porcentaje de varianza explicada por el primer factor es 26.69 y el ratio entre los dos primeros valores propios es 4.92. Mediante el análisis paralelo se obtuvieron 2 componentes. Cuando se analiza el VSS con complejidad 1 se obtiene un máximo de 0.794 con 1 componente, el gráfico que sigue muestra la solución para las diferentes complejidades. El criterio MAP da muy semejante con 1, 2 y 3 componentes. Al igual que en el caso del cuadernillo 1 se realizarán dos análisis.

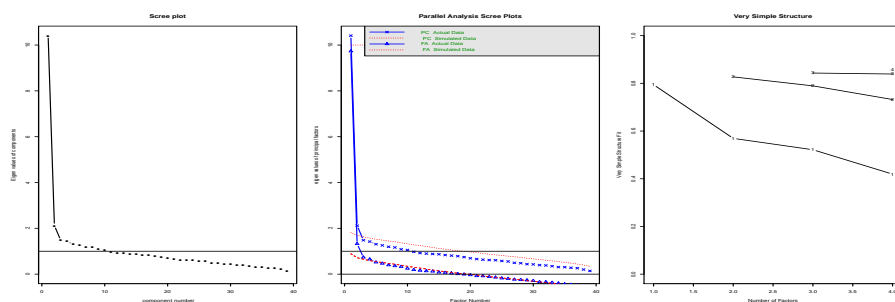


Figura 6.3: Prueba 2- Cantidad de factores

Para calibrar la prueba se utilizó el modelo logístico de 2 parámetros. Se estudió el ajuste al modelo tanto de las CCI como de las personas. También para el ajuste, se utilizó el procedimiento no paramétrico descrito en Luzardo y Rodríguez (2015) y luego se utilizaron tres distancias para determinar si las curvas paramétricas ajustaban.

Para medir el ajuste de las personas se usó el estadístico L_0 de Levine y Rubin (1979) y su versión estandarizada L_z propuesta por Drasgow et al. (1985). De este análisis se encontró que el 96.45 % de los sujetos ajustaron bien. Para el primer análisis de ajuste de los ítems se utilizó el Q_1 de Yen. Como este test tiene una distribución chi cuadrado el cual crece rápidamente y puede llevar a rechazar la hipótesis nula de ajuste, se decidió aplicar un método no paramétrico propuesto por Luzardo y Rodríguez (2015) el cual calcula el área entre las curvas y que llevó a aceptar el modelo 2P propuesto.

En los gráficos siguientes se pueden observar para el cuadernillo 1 las curvas estimadas mediante el modelo 2P y para los ítems 1, 5, 9 y 23 la curva estimada mediante el modelo 2P (en negro), mediante regresión no paramétrica (rojo) y polinómica (verde) así como la estimación isótónica para cada uno de los modelos anteriores (azul y celeste respectivamente). Luego, presentamos una gráfica para las curvas del cuadernillo 2.

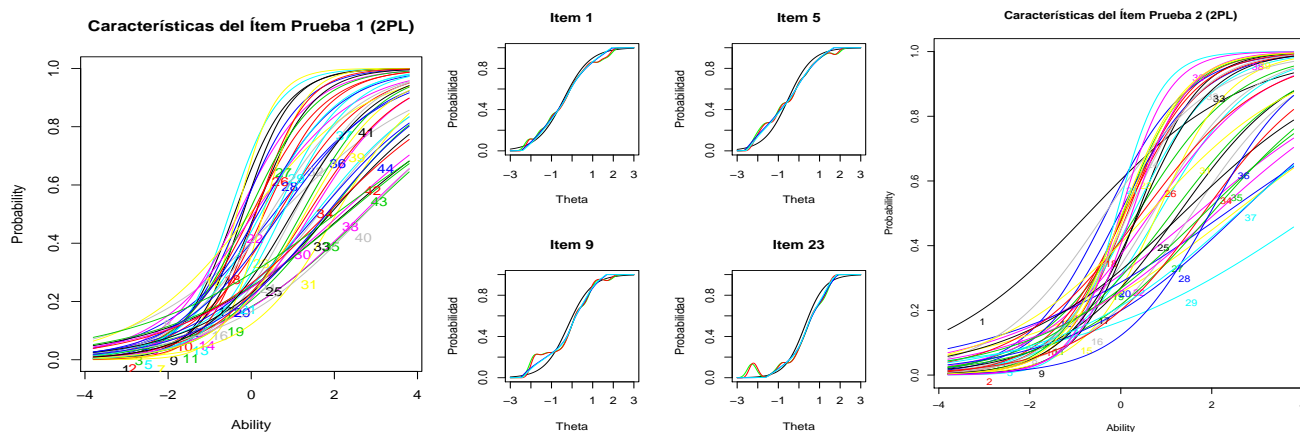


Figura 6.4: CCI cuadernillo 1

Figura 6.5: Items 1,5,9 y 13

Figura 6.6: CCI cuadernillo 2

Se encontró que el 98.46 % de los sujetos ajustaron bien. Las distancias entre las dos curvas para cada ítem se muestran en la tabla siguiente donde se ve que los valores son pequeños con respecto a los aceptados mediante simulación por lo que mantenemos el modelo 2P de los ítems.

Ítem	RECI-C1	RECE-C1	d3-C1	RECI-C2	RECE-C2	d3-C2
1	0.14	0.03	0.34	0.21	0.08	0.17
2	0.12	0.02	0.40	0.05	0.03	0.06
3	0.07	0.02	0.25	0.20	0.10	0.11
4	0.07	0.02	0.15	0.36	0.19	0.04
5	0.21	0.03	0.48	0.08	0.03	0.12
6	0.14	0.04	0.25	0.21	0.06	0.41
7	0.20	0.04	0.46	0.33	0.10	0.56
8	0.06	0.02	0.12	0.13	0.05	0.22
9	0.12	0.04	0.19	0.10	0.05	0.12
10	0.21	0.04	0.44	0.06	0.03	0.12
11	0.12	0.03	0.27	0.34	0.16	0.02
12	0.12	0.02	0.32	0.29	0.14	0.04
13	0.08	0.02	0.15	0.20	0.10	0.07
14	0.13	0.03	0.30	0.11	0.02	0.24
15	0.17	0.04	0.32	0.08	0.03	0.10
16	0.06	0.02	0.10	0.32	0.14	0.40
17	0.06	0.02	0.12	0.24	0.09	0.38
18	0.09	0.02	0.48	0.22	0.11	0.04
19	0.07	0.03	0.14	0.26	0.11	0.41
20	0.17	0.03	0.44	0.20	0.03	0.46
21	0.11	0.04	0.17	0.23	0.12	0.04
22	0.06	0.03	0.13	0.32	0.11	0.50
23	0.07	0.02	0.14	0.20	0.10	0.04
24	0.17	0.04	0.36	0.29	0.15	0.10
25	0.19	0.04	0.37	0.22	0.06	0.38
26	0.16	0.04	0.33	0.09	0.03	0.16
27	0.04	0.02	0.07	0.29	0.10	0.46
28	0.05	0.03	0.12	0.38	0.14	0.55
29	0.13	0.02	0.30	0.46	0.16	0.71
30	0.09	0.02	0.21	0.07	0.04	0.09
31	0.15	0.02	0.46	0.19	0.06	0.30
32	0.06	0.03	0.10	0.10	0.06	0.07
33	0.11	0.03	0.33	0.19	0.08	0.06
34	0.11	0.03	0.24	0.16	0.04	0.35
35	0.12	0.03	0.29	0.21	0.05	0.41
36	0.07	0.03	0.17	0.40	0.19	0.55
37	0.10	0.02	0.40	0.32	0.10	0.52
38	0.06	0.03	0.12	0.03	0.01	0.07
39	0.09	0.02	0.22	0.15	0.06	0.23
40	0.07	0.03	0.14			
41	0.05	0.02	0.13			
42	0.20	0.04	0.47			
43	0.04	0.02	0.06			
44	0.04	0.02	0.06			

Cuadro 6.1: Ajuste de los cuadernillos 1 y 2.

En este cuadro vemos como podemos usar la regresión no paramétrica para determinar si el modelo elegido es adecuado para los datos que se tienen.

A continuación se presentan los resultados correspondientes a la calibración de los ítems de las pruebas mediante el modelo de 2P (Rodríguez, 2016).

Ítem	b	error	a	error
1	-0.41	0.07	1.56	0.16
2	0.31	0.08	1.25	0.13
3	2.72	0.46	0.55	0.09
4	0.54	0.13	0.73	0.10
5	-0.53	0.06	1.71	0.18
6	-0.31	0.08	1.15	0.13
7	-0.35	0.05	2.20	0.22
8	0.68	0.17	0.57	0.09
9	-0.20	0.06	1.74	0.17
10	0.64	0.12	0.93	0.11
11	1.20	0.14	1.03	0.11
12	-0.24	0.07	1.37	0.14
13	0.58	0.09	1.22	0.12
14	2.66	0.44	0.57	0.10
15	-0.03	0.11	0.71	0.10
16	1.62	0.22	0.74	0.10
17	0.09	0.07	1.46	0.15
18	0.01	0.08	1.13	0.12
19	1.86	0.25	0.73	0.10
20	1.59	0.24	0.66	0.09
21	1.52	0.22	0.71	0.10
22	0.06	0.10	0.83	0.10
23	0.28	0.06	2.07	0.19
24	1.02	0.13	1.07	0.11
25	2.25	0.41	0.49	0.09
26	0.11	0.06	1.66	0.16
27	-0.08	0.07	1.21	0.13
28	0.15	0.09	1.00	0.11
29	0.23	0.09	1.07	0.12
30	1.44	0.17	0.92	0.11
31	2.04	0.22	0.97	0.12
32	0.45	0.11	0.91	0.11
33	1.89	0.28	0.64	0.09
34	1.36	0.17	0.89	0.11
35	2.28	0.43	0.48	0.09
36	0.70	0.13	0.80	0.10
37	0.75	0.10	1.17	0.12
38	2.14	0.38	0.52	0.09
39	1.29	0.15	1.00	0.11
40	2.73	0.43	0.59	0.10
41	0.97	0.13	0.98	0.11
42	1.81	0.30	0.57	0.09
43	2.03	0.43	0.43	0.09
44	1.81	0.25	0.70	0.10

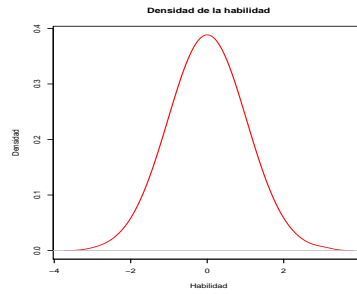
Ítem	b	error	a	error
1	-0.71	0.17	0.59	0.10
2	0.14	0.08	1.44	0.15
3	0.37	0.11	0.89	0.11
4	-0.32	0.09	1.11	0.13
5	0.84	0.13	0.96	0.11
6	1.68	0.34	0.47	0.09
7	2.48	0.52	0.43	0.09
8	0.70	0.12	0.94	0.11
10	0.37	0.08	1.50	0.15
12	0.07	0.08	1.28	0.13
13	0.04	0.08	1.39	0.14
14	-0.03	0.09	1.09	0.12
15	0.28	0.09	1.12	0.12
16	0.66	0.13	0.80	0.10
17	0.81	0.11	1.25	0.13
18	1.57	0.20	0.83	0.11
20	1.28	0.19	0.78	0.10
21	0.09	0.08	1.17	0.13
22	1.12	0.18	0.74	0.10
23	2.30	0.53	0.40	0.09
24	-0.09	0.06	2.05	0.20
25	2.08	0.39	0.50	0.09
26	0.10	0.07	1.48	0.15
27	-0.34	0.11	0.84	0.11
29	1.36	0.26	0.54	0.09
30	0.46	0.13	0.75	0.10
32	2.00	0.31	0.64	0.10
33	2.61	0.47	0.53	0.10
34	4.25	1.08	0.38	0.10
35	-0.01	0.06	1.87	0.18
36	0.62	0.16	0.61	0.09
37	0.17	0.08	1.24	0.13
38	0.40	0.09	1.25	0.13
39	1.81	0.24	0.77	0.10
40	1.84	0.33	0.55	0.09
41	1.96	0.21	1.01	0.12
42	2.58	0.48	0.51	0.09
43	0.24	0.07	1.62	0.16
44	0.03	0.07	1.45	0.15

Cuadro 6.3: Parámetros estimados Prueba 2

Cuadro 6.2: Parámetros estimados Prueba 1

6.4. Estimación no paramétrica en una dimensión

También se estimaron las CCI de los ítems mediante modelos no paramétricos. Se estimaron mediante el estimador unidimensional de regresión y mediante el estimador basado en polinomios locales de primer grado. Para esto se usó una ventana igual a 0.2 y un núcleo normal. Luego, a partir de estos estimadores, como es de suponer que las curvas deben ser crecientes al medir conocimiento, se estimaron las curvas isótonas unidimensionales, donde la ventana para la densidad se fijó en 0.15 y el núcleo nuevamente fue gaussiano. La densidad de los rasgos estimados de los sujetos está representada en la figura que sigue:



Se presenta a continuación la gráfica del ítem 13 del cuadernillo 1 en los cuatro modelos propuestos. En el Anexo se pueden ver las gráficas de los 44 ítems en las 4 modalidades estimadas.

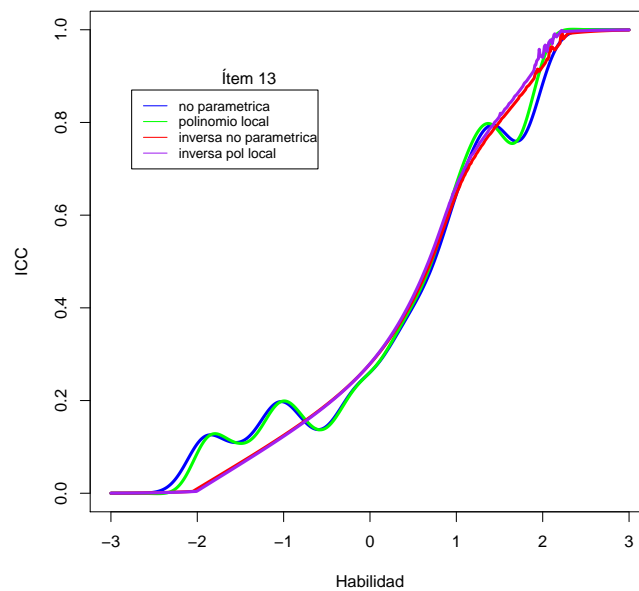


Figura 6.7: Estimaciones de la CCI ítem 13

6.5. Estimación no paramétrica en dos dimensiones

Se estimaron la densidad conjunta y las superficies de respuesta de los ítems usando los diferentes modelos. Primeramente se realizó un análisis factorial usando componentes principales los que se rotaron oblicuamente y, a partir de esto, se usaron las puntuaciones factoriales obtenidas para ordenar a los sujetos en cada dimensión a los efectos de estimar el rasgo. Se usó, como matriz de ancho de banda, una matriz diagonal con valores de 0.8 en la diagonal principal y el núcleo usado fue la distribución normal bivariada. El ancho de banda para el modelo isótono fue 0.15 y el núcleo gaussiano. La densidad conjunta estimada de los rasgos se muestra en la figura, siendo la correlación de 0.49. Las figuras siguientes muestran la densidad conjunta y las superficies de respuestas para el ítem 13 mediante los diferentes métodos. En el anexo se encontraran las superficies del cuadernillo 1 para todos los ítems.

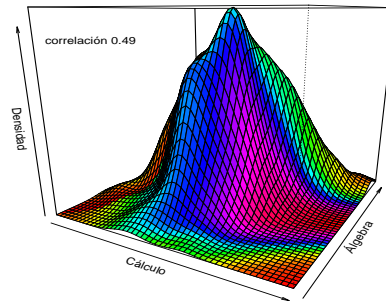


Figura 6.8: Densidad de la habilidad

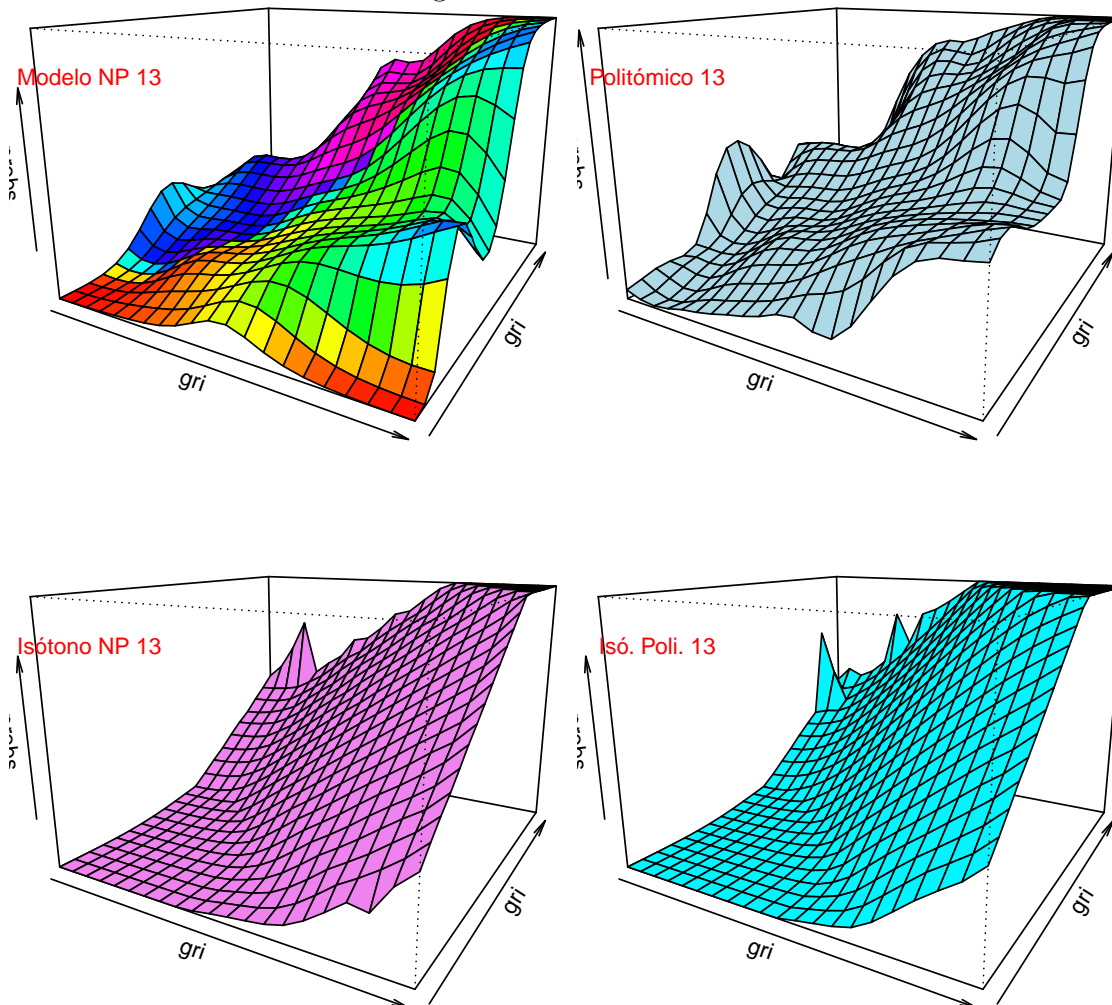


Figura 6.9: Superficies de respuesta estimadas

6.6. Aplicación al Children Depression Inventory

El Children Depression Inventory (CDI) es un test psicológico para el diagnóstico de la depresión en la infancia y adolescencia, de respuestas politómicas (0, 1 y 2), en el cual se miden cuatro rasgos en distintos ítems, tal como lo explica la siguiente tabla.

	Rasgo medido	Ítems
1	Autoestima Negativa Retraimiento	1, 4, 7, 10, 14, 20, 21, 22, 25
2	Autoestima Negativa Oposicionamiento	5, 9, 17, 26, 27
3	Ineficacia	2, 3, 13, 15, 23, 24
4	Ansiedad	6, 11, 19

El CDI presenta un coeficiente alfa de Cronbach que oscila entre 0.70 y 0.94 en distintos estudios realizados entre 1978 y 2003. La media es de 0.85, por lo que puede considerarse un instrumento fiable. También se ha encontrado que la estabilidad en el tiempo es muy alta para aplicarse a una población infantil y adolescente (del Barrio y Carrasco, 2003).

Este test fue aplicado a una muestra representativa de los estudiantes de Ciclo Básico de Educación Media de la ciudad de Montevideo. La muestra estuvo constituida por 605 adolescentes entre 12 y 13 años, de los cuales 324 eran de sexo femenino, 280 de sexo masculino y para 1 persona no existe este dato.

Como se tratan de respuestas graduadas se aplicó el modelo de regresión no paramétrica. Se encaró el problema desde dos ángulos: primero, simplemente aplicando el modelo de regresión y para ello se recodifican las variables de respuesta, considerando ahora variables indicatrices donde $Y_{ij} = 1$ $j = 0, 1, 2$, cuando el sujeto i elija la opción j . Estas variables dicotómicas son las que se usan en el estimador. El otro enfoque se basa en el modelo isótono. Para ello se crean otras variables auxiliares Z_{ij} la cual toma valor 1 si se ha elegido un puntaje menor o igual a j . Luego, $P(X_i = j) = P(X_i \leq j) - P(X_i < j)$.

A partir de esto se obtuvieron las funciones de respuesta. En las tres gráficas que siguen se ven las FRI para la dimensión Autoestima Negativa Retraimiento. Remitimos al lector al anexo para ver las gráficas de las otras dimensiones.

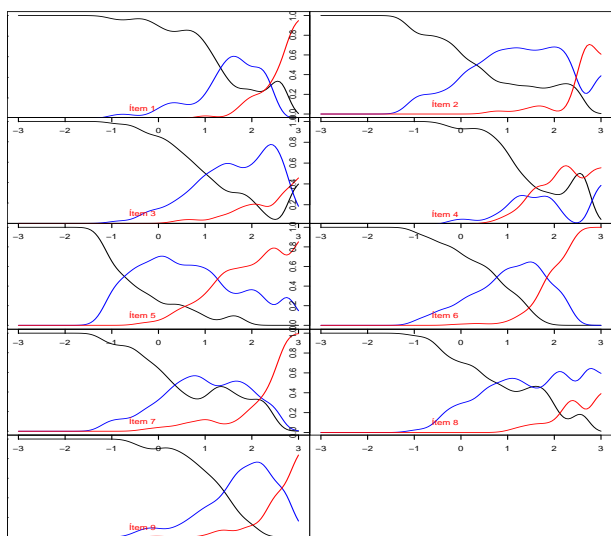


Figura 6.10: Autoestima Negativa Retraimiento Modelo no paramétrico

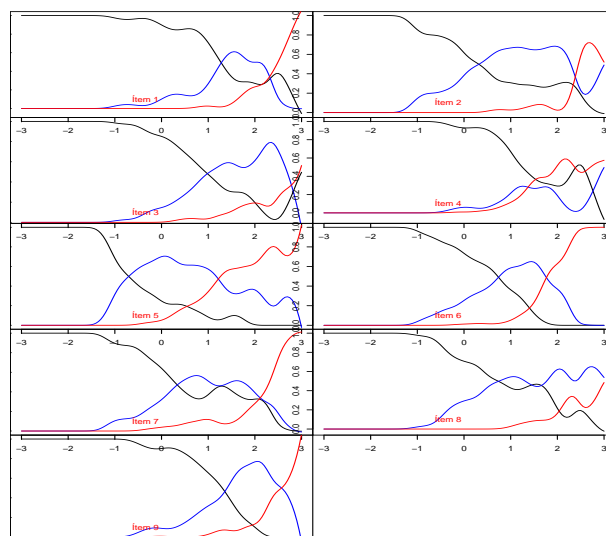


Figura 6.11: Autoestima Negativa Retraimiento Regresión polinómica

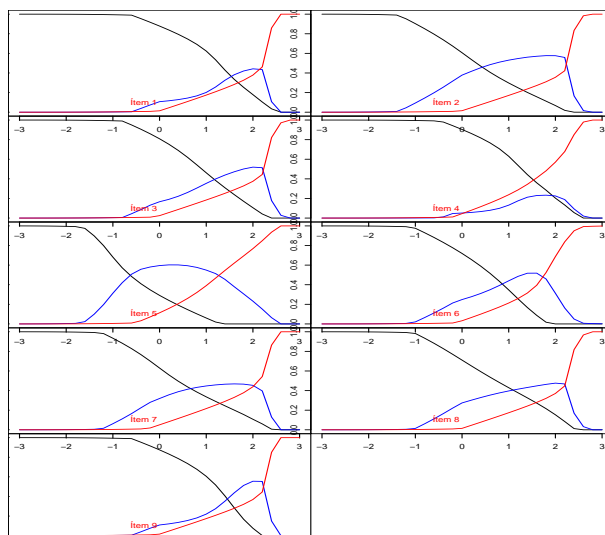


Figura 6.12: Autoestima Negativa Retraimiento modelo isótono

6.7. Estimación multidimensional

El CDI es un test de 4 dimensiones. Si nos apartamos del supuesto que los ítems dependen de una sola habilidad y pensamos que los diferentes constructos afectan a los diferentes ítems, se puede estimar este modelo de manera no paramétrica con 4 dimensiones. A los efectos de graficar y ver la relación entre las dimensiones, se proyectaron las curvas en los diferentes subespacios bidimensionales. Se graficaron, por lo tanto, las superficies de respuestas para las opciones 0,1 y 2 en los planos:

Retraimiento-Oposicionamiento

Retraimiento-Ineficacia

Retraimiento- Ansiedad

Oposicionamiento-Ineficacia

Oposicionamiento-Ansiedad

Ineficacia-Ansiedad

En el anexo se pueden ver las proyecciones bidimensionales para los diferentes planos y opciones de los ítems del CDI. Presentaremos aquí, a modo de ejemplo, las opciones del ítem 3 en el plano Retraimiento-Oposicionamiento.

Otra manera de ver la relación entre los diferentes constructos, es mantener dos de ellos constantes, hacer variar otro, tomar los valores del cuarto como variable independiente y graficar las probabilidades estimadas mediante el modelo de 4 dimensiones. Este tipo de gráfica es un símil a las curvas de nivel. En las siguientes gráficas se mantiene constante el nivel de Ineficacia y Ansiedad (en niveles predeterminados) en cada una, la variable independiente será el rasgo estimado en Autoestima Negativa Retraimiento y cada una de las curvas de color será un valor de la Autoestima Negativa Oposicionamiento, el rojo indica el nivel 3, el azul 2, el verde 1, amarillo 0, celeste -1, violeta -2 y gris -3.

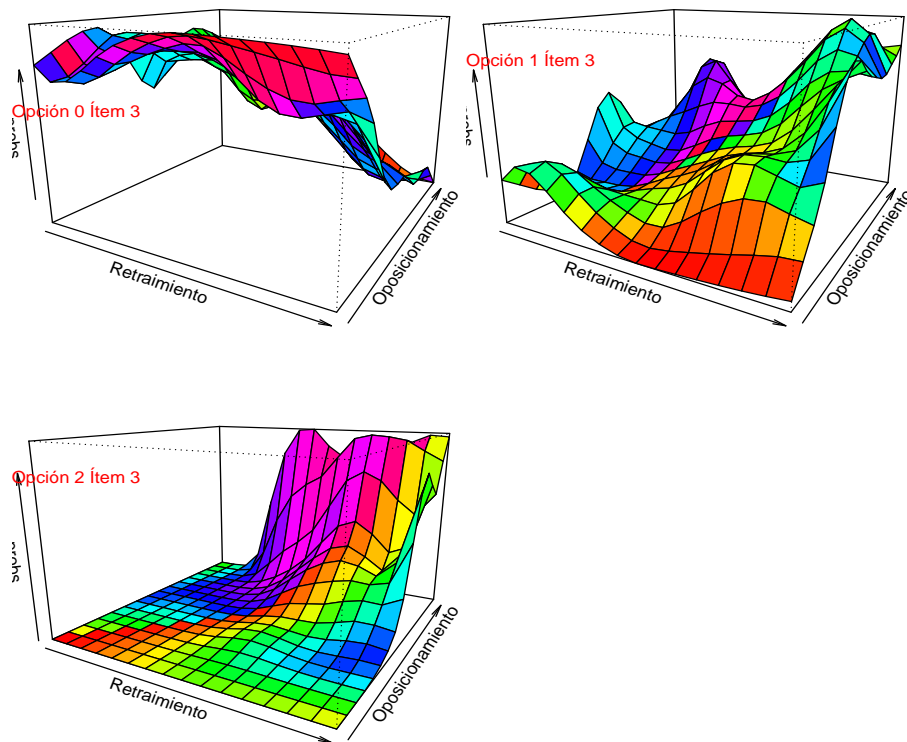


Figura 6.13: Proyección sobre el plano Retraimiento-Oposicionamiento Ítem 3

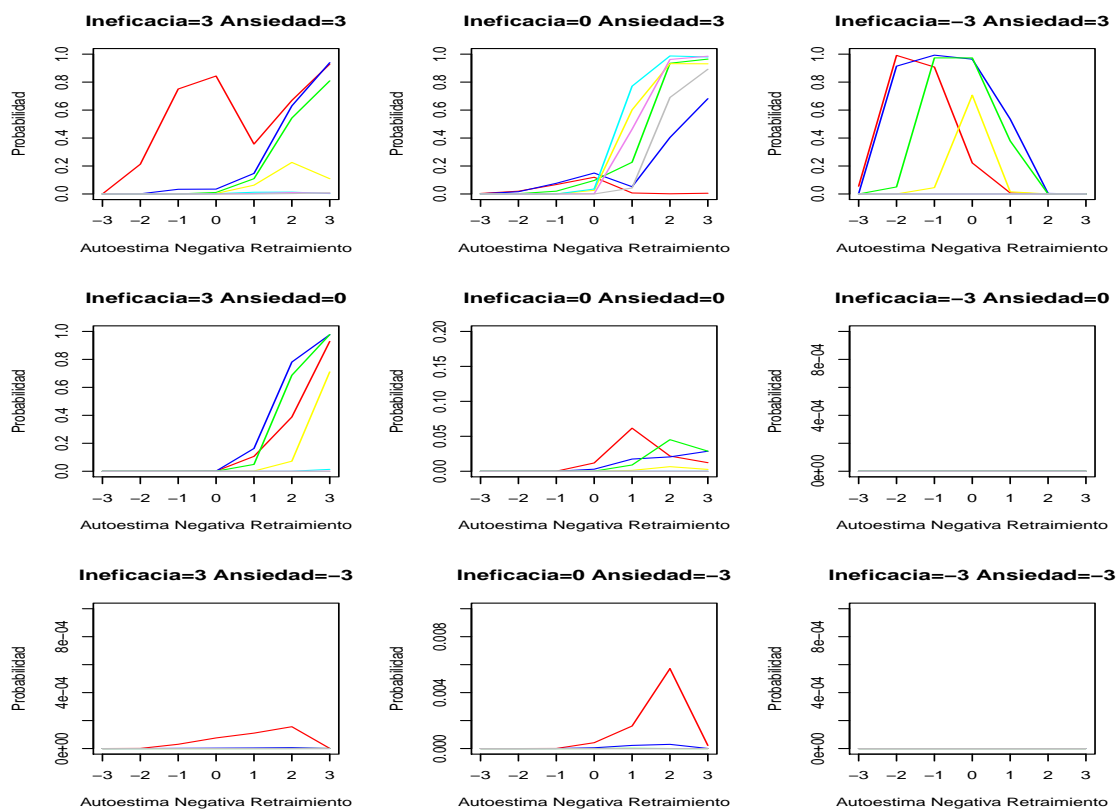


Figura 6.14: Curvas de nivel

Capítulo 7

Conclusiones

7.1. Aportes de la tesis

La tesis abarcó dos aspectos importantes de la TRI no paramétrica, el primero de ellos fue la estimación de modelos multidimensionales y el segundo, la estimación de modelos isótonos. Del primer aspecto se presentaron dos modelos no paramétricos multidimensionales para la estimación de los rasgos y de las probabilidades de respuesta condicionadas al nivel de rasgo. Estos modelos tanto pueden ser aplicados a ítems dicotómicos, politómicos o de respuesta continua. En el segundo aspecto, primero se presentó un modelo isótono unidimensional, y luego, un modelo multidimensional isótono en cada componente.

Como principal desarrollo de la tesis, además de los propios modelos, se tiene la demostración de la consistencia conjunta del estimador de rasgos y las CCI en el caso dicotómico, lo cual fue desarrollado detalladamente en el capítulo 3. Se comprobó mediante simulaciones la adecuación de estos estimadores calculando varias medidas de ajuste y comparándolos con la estimación paramétrica de los modelos. Surgió de este análisis que los modelos no paramétricos ajustan muy bien y que su ajuste es comparable a la estimación paramétrica en el caso que los modelos sigan las ecuaciones funcionales adecuadas a los modelos y son mejores cuando los ítems siguen otros modelos.

Se estudió la influencia de los diferentes núcleos y ventanas en los estimadores, verificándose que estos son buenos ante diferentes núcleos y son influenciados ante variaciones de las ventanas. Se presentaron métodos de elección de la ventana para ayudar en este problema. Se analizó también la varianza del estimador del rasgo, mediante simulación.

Además, se obtuvieron otros resultados de menor importancia y con limitaciones, ya que se simplificó el problema para que pudiera ser abordado mediante métodos usuales de regresión no paramétrica. Estos pueden ser una buena aproximación porque aún carecemos de la distribución (exacta o asintótica del estimador del rasgo). Se presentó, entonces, un estimador de la densidad conjunta en el caso de que el rasgo sea medido sin error. Se demostró que es consistente y se obtiene su distribución asintótica. A partir de esto, sugerimos posibles aproximaciones para el caso general, para la esperanza, varianza y distribución.

Para las CCI, también considerando el rasgo medido sin error, se probó la consistencia y distribución asintótica. También se encuentra la distribución asintótica del estimador isótono en este caso.

Se realizaron 17 estudios de simulación para analizar distintos elementos que influyen en la performance de los estimadores.

Finalmente, se aplicaron los modelos a dos conjuntos de datos reales: la prueba de evaluación

de aprendizajes de matemática al ingreso de la Universidad de la República y al test CDI (Children Depression Inventory). En la prueba de matemática se usaron los modelos unidimensionales y bidimensionales para calibrar los ítems. En el CDI se aplicaron los modelos unidimensionales (no isótono e isótono) para calcular las funciones de respuesta a ítems de respuesta graduada. Luego, se estimó un modelo multidimensional de cuatro dimensiones y se mostraron las proyecciones de este modelo sobre los espacios bidimensionales determinados por los diferentes constructos del CDI.

Surge, en primer lugar, como un hecho interesante sobre los resultados y aportes mencionados, que en el caso multidimensional no fue necesario conocer la distribución conjunta de los rasgos, sino que solamente se pide el conocimiento de las marginales o tomar sin pérdida de generalidad estas marginales uniformes. Se usó fuertemente la no identificabilidad de la escala y la reconstrucción de la distribución original a través de las cópulas para reducir el estudio a marginales uniformes. Este hecho es importante, pues, en general, se ignora la relación de dependencia entre los rasgos. Para estimar esta dependencia, la densidad conjunta y la cópula se estiman a partir de las marginales. Creemos que esto es un aporte relevante, dado que no es necesario presuponer conocimiento a priori del rasgo.

Destaca, a nuestro juicio, la simplicidad de los modelos, su flexibilidad ante diferentes tipos de curvas, su adecuación a diferentes tipos de ítems y lo fácil de implementar su estimación. Basta observar las simulaciones para justificar estas afirmaciones. En particular, en el capítulo 1 de la tesis, desarrollamos los algoritmos de estimación de los modelos paramétricos, que son iterativos y con una velocidad lineal en lograr la convergencia, al contrario de los estimadores no paramétricos, que no son iterativos y de sencilla programación.

La simplicidad de los modelos se deriva de que, para la estimación del rasgo solo se utiliza la función de distribución empírica, y las estimaciones de las CCI resultan de métodos de estimación no paramétrica de curvas de regresión basados en núcleos. Se considera un rasgo en R^d , $\Theta = (\Theta_1, \dots, \Theta_d)$ y para estimarlo sencillamente se consideraron tantas funciones de ordenación como rasgos, que dependen de un solo rasgo y luego en cada dimensión se toma el cuartil empírico. Respecto a la estimación de las CCI el primer enfoque está basado en el estimador de Nadaraya Watson, que es simplemente un promedio ponderado local. Toma la forma:

$$(7.1) \quad \widehat{P_j(\theta)} = \frac{\sum_{i=1}^N X_{i,j} K_H(\widehat{\theta}_i - \theta)}{\sum_{i=1}^N K_H(\widehat{\theta}_i - \theta)}$$

El segundo método está basado en polinomios locales de grado 1 los cuales son de:

$$(7.2) \quad \min_{\beta_0, \beta_1} \sum_{i=1}^N (X_{i,j} - \beta_0 - (\widehat{\theta}_i - \theta)^t \beta_1)^2 K_H(\widehat{\theta}_i - \theta),$$

que tiene como solución $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X$ donde

$$\Sigma = \begin{pmatrix} 1 & (\widehat{\theta}_1 - \theta)^t \\ \vdots & \vdots \\ 1 & (\widehat{\theta}_N - \theta)^t \end{pmatrix}, X = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{N,j} \end{pmatrix}$$

$$W = \begin{pmatrix} K_H(\widehat{\theta}_1 - \theta) & 0 & \dots & 0 \\ 0 & K_H(\widehat{\theta}_2 - \theta) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \dots & K_H(\widehat{\theta}_N - \theta) \end{pmatrix}$$

El estimador de la CCI es β_0 y el estimador del gradiente es β_1 . Este se puede calcular fácilmente usando solo operaciones básicas de matrices mediante,

$$\widehat{P}_j^1(\theta) = e_1^t(\widehat{\beta}_0, \widehat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X$$

donde e_1 es el primer vector de la base canónica.

El aspecto más técnico de la tesis es la obtención de las condiciones para lograr la consistencia conjunta de los estimadores del rasgo y las CCI cuando el número de ítems y de sujetos crecen conjuntamente. En esta parte, adicionalmente, se definieron ciertos objetos matemáticos como son la familia de las funciones asintóticamente Lipchitz y se caracterizaron, lo que constituye otro elemento de interés.

Se encontraron 12 hipótesis necesarias que dieron origen a 4 resultados fundamentales. Las hipótesis son esencialmente sencillas, pero destaca el hecho que el crecimiento conjunto de la cantidad de ítems y de sujetos no es arbitrario sino que está controlado por un polinomio. Las hipótesis planteadas son:

- 1) Independencia local.
- 2) El rasgo Θ tiene distribuciones marginales uniformes $U[0, 1]$.
- 3) Para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$, tales que para $1 \leq l \leq d$:

$$(7.3) \quad \sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2$$

- 4) Las $F_{n,l}$ correspondientes son asintóticamente Liptchitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$.
- 5) $\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0$.
- 6) Para $0 < \alpha < 1/2$. $\frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$
- 7) $H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}} < N < M n^r$ con $r > d + 1$.
- 8) $\lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$.
- 9) El núcleo K es simétrico, acotado y con soporte compacto.
- 10) Las CCI tienen derivadas parciales continuas en $(0, 1)^d$ y, además, en cada compacto $[a, b]^d \subset (0, 1)^d$ y todo par (n, i) están uniformemente acotadas.
- 11) Para $0 < \alpha < 1/2$ $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$.
- 12) El núcleo K es Lipchitz de constante J .

Como afirmamos anteriormente, obtuvimos cuatro resultados importantes, la consistencia de la estimación de los rasgos, la consistencia de la estimación de las CCI, una desigualdad para el ECM y la consistencia conjunta.

Resultado 1 Si se cumplen las hipótesis 1) a 7) $\max_{k=1, \dots, N} \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad.

Si además se cumple 8), entonces $\max_{k=1, \dots, N} \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente.

Resultado 2 Dadas las hipótesis 1) a 12) y para $\theta_0 \in (0, 1)^d$ fijo

$$\max\{ECM(\widehat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

Resultado 3 Dadas las hipótesis 1) a 12) $\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\widehat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$ si $n \rightarrow +\infty$ completamente con $(a, b)^d \subset (0, 1)^d$.

Resultado 4 Si se cumplen las hipótesis 1) a 12) si $n \rightarrow +\infty$ con probabilidad 1 se cumple

$$\max\{\max_{k=1, \dots, N} \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\widehat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0$$

Otro aspecto fundamental que aborda la tesis son los modelos isótonos. El estimador isótono planteado mantiene la simplicidad y la condición de no ser iterativo. Utiliza las estimaciones no paramétricas anteriores y luego un estimador de la densidad, el cual integra. Se destaca, además,

la elegancia de su justificación y el hecho de dar curvas suaves, al contrario de sus competidores. Estima la inversa de la CCI mediante:

$$(7.4) \quad \widehat{P}^{-1}(\theta) = \frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}) - u}{h_d}\right) du.$$

Finalmente, el estimador de \widehat{P} se obtiene por reflexión de \widehat{P}^{-1} con respecto a $y = x$.

Su extensión al caso multidimensional también es muy sencilla, como lo muestra el algoritmo de cinco pasos al que da origen que se puede leer en el capítulo 4.

Este trabajo también aporta otros resultados más limitados, que pueden ser vistos como casos sencillos o posibles aproximaciones. Ante el problema de estimar la densidad conjunta de los rasgos, decidimos utilizar el estimador de densidad del rasgo estimado. Es claro que incurrimos en un sesgo y en una subestimación de la varianza del estimador, pero nos permite fácilmente obtener intervalos de confianza aproximados. Es claro que estos resultados son de un carácter limitado, pero tienen la ventaja que se obtienen fácilmente a partir de la teoría clásica de estimación no paramétrica de densidades. La forma general del estimador de la densidad del rasgo multidimensional propuesta es:

$$\hat{f}_{N,H}(\theta) = \frac{1}{N} \sum_{i=1}^N K_H(\theta - \widehat{\Theta}_i) \quad \text{donde} \quad K_H = |H|^{-1} K(H^{-1}\theta)$$

donde H es una matriz $d \times d$ simétrica definida positiva llamada matriz de ancho de ventanas o de banda y K es un núcleo de dimensión d . En base a este estimador, como $\max_{k=1,\dots,N} \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$, o para el caso de medir las variables sin error, podemos pensar en aproximaciones del valor esperado de la densidad en $\theta \in R^d$ por:

$$E(f_{N,H}(\theta)) \approx f(\theta) + \frac{1}{2}\mu_2(K)tr(H^t \mathfrak{H}_f(\theta)H)$$

$$V(f_{N,H}) \approx \frac{1}{N|H|} \|K\|_2^2 f(\theta)$$

$$\text{y que } \sqrt{N|H|} \left(\widehat{f}_N(\theta) - f(\theta) \right) \rightarrow N(0, f(\theta) \|K\|_2^2).$$

Se probó, además, que si las variables son medidas sin error, o sea, $\widehat{\Theta}_i = \Theta_i$ para todo i , el estimador de Nadaraya Watson cumple que,

$$E(\widehat{P}(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \mu_2(K) \frac{\nabla_P(\theta)^t H H^t \nabla_f(\theta)}{f(\theta)} + \frac{1}{2}\mu_2(K)tr(H^t W_P(\theta)H)$$

$$V(\widehat{P}(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

y el estimador basado en polinomios locales cumple

$$E(\widehat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \frac{1}{2}\mu_2(K)tr(H^t W_P(\theta)H)$$

$$V(\widehat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

Finalmente, bajo condiciones generales, se establece un resultado para la distribución asintótica del estimador de las CCI probando que converge débilmente a la normal y lo mismo ocurre para el modelo isótono.

Creemos entonces que los modelos presentados en esta tesis constituyen un aporte a la TRI noparamétrica basada en regresión con núcleos. Los modelos resultaron ser consistentes, flexibles y funcionaron muy bien en las simulaciones y cuando se aplicó a datos reales. Son robustos ante diferentes núcleos y distribuciones; en dimensiones bajas el tamaño muestral necesario para obtener un buen ajuste es moderado. El mayor cuidado debe tenerse en la elección de la ventana. La implementación de los algoritmos fue sencilla comparados con sus competidores paramétricos. Igualmente el método tiene limitaciones las cuales en la sección que sigue se harán explícitas, constituyéndose estas en posibles desarrollos futuros.

7.2. Limitaciones y problemas abiertos

Los modelos encontrados tienen varias limitaciones y quedan planteados algunos problemas a resolver. Un primer problema que ha estado presente es no tener la distribución de los estimadores de los rasgos, o su distribución asintótica. Si bien esto no fue necesario para probar la consistencia conjunta, sí lo es para poder encontrar intervalos de confianza para los rasgos, el cálculo de las varianzas de los estimadores de las CCI y la distribución asintótica de estos. Diremos al respecto que es un problema de difícil solución. No se abarca este tema en este trabajo, pero como punto de partida para poder encontrar la solución a este problema está el excelente artículo de Sperlich (2008), donde encuentra la distribución del estimador de densidad cuando la variable es medida con error si se cumplen ciertas condiciones. No es inmediata su aplicación, pues en ese caso debe conocerse el sesgo, su derivada parcial y la varianza del error, además de otras condiciones. Sin embargo, puede intentar estimarse mediante otros métodos, por ejemplo, bootstrap. Queda por resolver este problema que, a nuestro juicio, es interesante por los resultados y por la propia dificultad que plantea.

Otra limitante es que este método requiere que las curvas se calibren en una muestra representativa de la población, que hasta filosóficamente es un problema al depender de la población en que se prueba.

Surge también el problema de la dimensionalidad, donde para dimensiones superiores, la cantidad de datos necesarios para una correcta estimación crece mucho con la dimensión. Por ejemplo, en el caso de querer estimar la densidad del rasgo, si suponemos que este es gaussiano, estimamos la densidad mediante un núcleo normal y se usa la ventana óptima para tener un error cuadrático medio menor a 0.01, entonces en el caso de dimensión 4, necesitamos 223 sujetos, pero si pasamos a dimensión 7, ya necesitamos 10.700 y en el caso de dimensión 10, son necesarios 842.000 sujetos. Igualmente, para los instrumentos clásicos que tienen hasta 5 dimensiones solo necesitamos 768 sujetos.

Las funciones utilizadas para ordenar a los sujetos a los efectos de estimar los rasgos es otro problema a resolver. Un hecho que juega un rol importante en la demostración de la consistencia es que las funciones sean independientes de los rasgos que no estiman. En la práctica resulta que muchas de estas funciones dependen principalmente de un rasgo, pero no son independientes de los otros. Queda por encontrar qué tipo de dependencia débil sobre los otros rasgos permite la consistencia de los estimadores.

Estrictamente hablando, deberíamos trabajar solo con núcleos con soporte compacto, lo cual excluye, por ejemplo, al núcleo normal. Cuando lo utilizamos vemos que se obtienen buenos ajustes. Esto nos indica que deberíamos poder probar la consistencia para una clase de núcleos más amplia. Consideramos, a partir de los resultados del capítulo 3, que las condiciones necesarias sobre el núcleo pueden ser del estilo:

a) Existen $C_1, C_2 > 0$ y H tales que $C_1 H(\|\Theta\|) \leq K(\theta) \leq C_2 H(\|\Theta\|)$ con H acotada, decreciente y tal que $t^d H(t) \rightarrow 0$ si $|t| \rightarrow \infty$.

b) Existen $c, r > 0$ tal que $K(\theta) \geq c \mathbb{1}_{\{\|\Theta\| \leq r\}}(\theta) \chi_{\|\theta\| \leq r}(\theta)$.

Bibliografía

- [1] Abad, F., Olea, J. y Ponsoda, V. (2009). The Multiple-Choice Model Some Solutions for Estimation of Parameters in the Presence of Omitted Responses. *Applied Psychological Measurement*, 33(3), 200-221.
- [2] Ackerman, T. A. (1992): A didactic explanation of item bias, item impact, and item validity from multidimensional perspective, *Journal of Educational Measurement*, 29(1), 67-91.
- [3] Akaike, H. (1956): On a 0-1 process and some of his applications, *Annals of the Institute of Statistical Mathematics*, 8, 87-94.
- [4] Andersen, E. B. (1973): Conditional inference for multiple-choice questionnaires, *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- [5] Andersen, E. B. (1977): Sufficient statistics and latent trait models, *Psychometrika*, 42, 69-81.
- [6] Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- [7] Andrich, D. (1988). The application of an unfolding model of the PIRT type for the measurement of attitude. *Applied Psychological Measurement*, 12, 33-51.
- [8] Andrich, D. (2005). Georg Rasch: Mathematician and statistician. En: Kempf-Leonard, K. (Ed.), *Encyclopedia of Social Measurement* (Volume 3) (pp. 299-306). Amsterdam, The Netherlands: Academic Press.
- [9] Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch rating model which dispels any 'threshold disorder controversy'. *Educational and Psychological Measurement*, 73, 78-124.
- [10] Andrich, D. y Luo, G. (2003). Conditional estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement*, 4, 205-221.
- [11] Andrich, D., Sheridan, B. S. and Luo, G. (2014). *RUMM2030: Rasch Unidimensional Models for Measurement*. Perth, Western Australia: RUMM Laboratory.
- [12] Baker, F. (1992): *Item Response Theory Parameter Estimation Techniques*, New York: Marcel Dekker, Inc.
- [13] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions: the theory and application of isotonic regression*. New York: Wiley.
- [14] Barton, M.A. y Lord, F. M. (1981): An upper asymptote for the three parameter logistic item response model, *Research Bulletin*, Princeton, Nueva Jersey: Educational Testing Service.
- [15] Bartholomew, D.J. (1987): *Latent Variable Models and Factor Analysis*, London: Charles Griffin and Company.
- [16] Bartholomew, D. J. (1995). Spearman and the origin and development of factor analysis. *British Journal of Mathematical and Statistical Psychology*, 48, 211-220.
- [17] Bejar, I. (2010). Recent developments and prospects in item generation. En S. E. Embretson (Ed.), *Measuring Psychological Constructs: Advances in Model-Based Approaches* (pp. 201-226). Washington, DC: American Psychological Association Books.
- [18] Berkson, J. (1955): Maximum likelihood and minimum chi square estimates of the logistic function, *Journal of the American Statistical Association*, 50, 120-162.
- [19] Billingsley, P. (1968): *Convergence of Probability Measures*, New York: Wiley.
- [20] Birnbaum, A. (1968): Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord y M.R. Novick (eds.), *Statistical Theories of Mental Test Scores*, pp.397-472, Reading, MA: Addison-Wesley.
- [21] Bock, R. D. y Aitkin, M. (1981): Marginal maximum likelihood estimation of item parameters. An application of an EM algorithm. *Psychometrika*, 46, 443-459.

- [22] Bock, R. D. y Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- [23] Bock, R. D. y Lieberman, M. (1970): Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- [24] Bock, R. D. (1972): Estimating multinomial response relations. En R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao, y K.J.C. Smith (eds.), *Essays in Probability and Statistics*, Chapel Hill, NC, The University of North Carolina Press.
- [25] Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, 16, 21-33.
- [26] Bolt, D., y Johnson, T. (2009). Applications of a MIRT model to self-report measures: Addressing score bias and DIF due to individual differences in response style. *Applied Psychological Measurement*, 33, 335-352.
- [27] Boucheron, S., Lugosi, G. y Massart, P.(2003): Concentration inequalities using the entropy method. *Annals of Probability* 31, 1583-1614.
- [28] Bouwmeester, S., Sijtsma, K., y Vermunt, J.K. (2004). Latent class regression analysis for describing cognitive developmental phenomena. An application to transitive reasoning. *European Journal of Developmental Psychology*, 1, 67-86.
- [29] Bradlow, E. T., Wainer, H., y Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153-168.
- [30] Brunk, H.D. (1955). Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics*, 26, 607-616.
- [31] Cai, L., Choi, K., Hansen, M. y Harrell, L. (2016). Item Response Theory. *Annual Review of Statistics and Its Applications* 3, 297-321.
- [32] Carlson, J. E., y Davier, M. (2013). *Item Response Theory*. ETS Research Report Series, 2013(2), i-69.
- [33] Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245-276.
- [34] Cheng, K. F. & Lin, P. E. (1981). Nonparametric estimation of a regression function. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2), 223-233.
- [35] Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- [36] Coombs, C.H. (1964). *A theory of data*. Ann Arbor, MI: Mathesis Press.
- [37] Courant, R. y Hilbert, D. (1953). *Methods of Mathematical Physics* (Volume I). New York: Interscience Publishers.
- [38] Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4), 377-403.
- [39] Debeer, D., y Janssen, R. (2012). Modeling Item-Position Effects within an IRT Framework. *Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME)*, Vancouver.
- [40] De Boeck, P. (2008). Random item IRT models. *Psychometrika*, 73, 533-559.
- [41] del Barrio, V. y Carrasco, M.A. (2003). *Adaptación española del Children Depression Inventory*, Madrid: UNED.
- [42] De Koning, E., Sijtsma, K., y Hamers, J. H. (2002). Comparison of four IRT models when analyzing two tests for inductive reasoning. *Applied Psychological Measurement*, 26(3), 302-320.
- [43] Del Barrio, V., Moreno-Rosset, C., y López-Martínez, R. (1999). El Children's Depression Inventory,(CDI; Kovacs, 1992). Su aplicación en población española. *Clínica y Salud*, 10(3), 393-416.
- [44] Delecroix, M., & Thomas-Agnan, C. (2000). Spline and kernel regression under shape restrictions. En M.G. Schimek (Ed.) *Smoothing and Regression: Approaches, Computation, and Application*, pp.109-133. New York: Wiley.
- [45] Dempster, A. P., Laird, N. M. y Rubin, D. B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the royal Statistical Society, Series B*, 39, 1-38.
- [46] Dette, H, Neumeyer, N. y Pilz, K. (2006): A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3), 469-490.
- [47] Devroye, L. y Györfi, L. (1985): *Nonparametric Density Estimation, the L1 view*, New York: Wiley.
- [48] Dorans, N. J. y Holland, W. P. (1993): DIF detection and description: Mantel-Haenszel and standardization. En W. P. Holland y H. Wainer (eds.), *Differential Item Functioning* pp. 35-66) Hillsdale, NJ: LEA.
- [49] Douglas, J. (1997): Joint Consistency of Nonparametric Item Characteristic Curve and Ability Estimation. *Psychometrika*, 62, 7-28.
- [50] Douglas, J. y Cohen, A (2001): Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234-243.

- [51] Douglas, J., Kim, H.R., Habing, B. y Gao, F. (1998): Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, 23, 129-151.
- [52] Drasgow, F., Levine, M.V., Williams, B., McLaughlin, M.E., y Candell, G.L. (1989): Modeling incorrect responses to multiple-choice item with multilinear formula score theory. *Applied Psychological Measurement*, 13, 285-299.
- [53] Draney, K., and Wilson, M. (2008). A LLTM approach to the examination of teachers' ratings of classroom assessment tasks. *Psychology Science Quarterly*, 50, 417-432.
- [54] Edwards, A. L., and Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 17, 169-180.
- [55] Ellis, J. L. y Junker, B. w. (1997): Tail-measurability in monotone latent variable models, *Psychometrika*, 62, 495-523.
- [56] Elliot, C.D., Murray, D.J. y Saunders, R. (1977): *Goodness of fit to the Rasch as a criterion of test unidimensionality*. Manchester: University of Manchester.
- [57] Embretson, S. E. (1984). A general multicomponent latent trait model for response processes. *Psychometrika*, 49, 175-186.
- [58] Embretson, S.E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56, 495-515.
- [59] Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, 64, 407-433.
- [60] Embretson, S. E. and Yang, X. (2013). A multicomponent latent trait model for diagnosis. *Psychometrika*, 78, 14-36.
- [61] Embretson, S. E. (2016) Multicomponent Models. En En: W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume one: Models, pp.225-241. Boca Raton, FL: Chapman & Hall/CRC Press.
- [62] Fan, J. y Truong, Y.K.(1993). Nonparametric regression with errors in variables. *Annals of Statistics*, 21, 1900-1925.
- [63] Fan, J., & Gijbels, I. (1996). Local polynomial modelling and its applications: monographs on statistics and applied probability 66 (Vol. 66). CRC Press.
- [64] Feller, W.(1978). *An introduction to probability theory and its applications*, New York: Wiley.
- [65] Ferrando, P.J.(2004): Kernel-smoothing estimation of item characteristic functions for continuous personality items: an empirical comparison with the linear and the continuous-response models *Applied Psychological Measurement*, 28 (2), 95-109.
- [66] Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3-26.
- [67] Fisher, E.(1992): A Skorohod representation and an invariance principle for sums of weighted iid random variables, *Journal of Mathematics*, 22, 169-179.
- [68] Fischer, G. H. (1995a). Linear logistic models for change. En: G. H. Fischer and I. W. Molenaar (Eds.) *Rasch Models: Foundations, Recent Developments, and Applications*, pp. 157-180. New York: Springer.
- [69] Fischer, G. H. (1995b). The linear logistic test model. En: G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications*, pp. 131-155. New York: Springer.
- [70] Fidalgo, A. M. (1996): Funcionamiento diferencial de los ítems. Procedimiento Mantel-Haenzel y modelos loglineales. Tesis doctoral no publicada. Universidad de Oviedo.
- [71] Fidalgo, A.M. (1994): MHDIF: A computer program for detecting uniform and nonuniform differential item functioning with Mantel- Haenzel procedure, *Applied Psychological Measurement*, 18(3), 300.
- [72] Fischer, G. H. y Scheiblechner, H. H. (1970): Algorithmen und programmen für das probabilistische testmodel von Rasch, *Psychologische Beitrage* 12, 23-51.
- [73] Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37, 359-374.
- [74] Fischer, G. H. (1995a). Linear logistic models for change. En: G. H. Fischer and I. W. Molenaar (Eds.) *Rasch Models: Foundations, Recent Developments, and Applications*, (pp. 157-180). New York: Springer.
- [75] Fischer, G. H. (1995b). The linear logistic test model. En: G. H. Fischer and I. W. Molenaar (Eds.), *Rasch Models: Foundations, Recent Developments, and Applications*, pp. 131-155. New York: Springer.
- [76] Fitzpatrick, A. R., Link, V. B., Yen, W. M., Burket, G. R., Ito, K., and Sykes, R. C. (1996). Scaling performance assessments: A comparison of one-parameter and two-parameter partial credit models. *Journal of Educational Measurement*, 33, 291-314.
- [77] Fraser, C., y McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.

- [78] Friedman, J. & Tibshirani, R. (1984). The monotone smoothing of scatterplots. *Technometrics*, 26(3), 243-250.
- [79] García, P.E.; Abad, F.J.; Olea, J. y Aguado, D. (2013). A new IRT-based standard setting method: Application to elath-Listening. *Psicothema*, 25 (2), 238-244.
- [80] García, P.E., Olea, V. y de la Torre, J. (2014). Application of Cognitive Diagnosis Models to Competency-Based Situational Judgment Tests. *Psicothema*, 26 (3), 372-377.
- [81] Garwood, F. (1941): The application of maximum likelihood to dosage-mortality curves, *Biometrika*, 32, 46-58.
- [82] Geyer, C. 2011. Introduction to MCMC. In S. Brooks, A. Gelman, G. Jones, and X. Meng (Eds), *Handbook of Markov Chain Monte Carlo* (p. 3748). Boca Raton, Florida: CRC.
- [83] Giampaglia, G. (1990). *Lo scaling unidimensionale nella ricerca sociale*. Napoli: Liguori Editore.
- [84] Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 24, 229-252.
- [85] Gijbels, I. (2005). Monotone regression. In N. Balakrishnan, S. Kotz, C.B. Read and B. Vadačovic (Eds.) *The Encyclopedia of Statistical Sciences*, 2nd. edition. Hoboken, New York: Wiley.
- [86] Gierl, M. and Haladyna, T. (2012). *Automatic Item Generation*. New York: Taylor and Francis: Rutledge.
- [87] Grayson, D.A. (1988): Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika* 53, 383-392.
- [88] Grebliki, W., Krzyzak, A., Pawlak, M. (1984): Distribution-free pointwise consistency of kernel regression estimate. *The Annals of Statistics*, 12, 1570-1575.
- [89] Green, B. F., Jr. (1950a). *A general solution for the latent class model of latent structure analysis* (Research Bulletin No. RB-50-38). Princeton, NJ: Educational Testing Service.
- [90] Green, B. F., Jr. (1950b). *A proposal for a comparative study of the measurement of attitude* (Research Memorandum No. RM-50-20). Princeton, NJ: Educational Testing Service.
- [91] Green, B. F., Jr. (1950c). *A proposal for an empirical evaluation of the latent class model of latent structure analysis* (Research Memorandum No. RM-50-26). Princeton, NJ: Educational Testing Service.
- [92] Green, B. F., Jr. (1950d). *Latent structure analysis and its relation to factor analysis* (Research Bulletin No. RB-50-65). Princeton, NJ: Educational Testing Service.
- [93] Green, B. F., Jr. (1951a). A general solution for the latent class model of latent structure analysis. *Psychometrika*, 16, 151-166.
- [94] Green, B. F., Jr. (1952). Latent structure analysis and its relation to factor analysis. *Journal of the American Statistical Association*, 47, 71-76.
- [95] Green, B. R., Jr. (1980). *Ledyard R Tucker's affair with psychometrics: The first 45 years*. Paper presented at a special symposium in honor of Ledyard R Tucker, Champaign, IL: The University of Illinois.
- [96] Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- [97] Guttman, L. (1950). The basis for scalogram analysis. En S.A. Stouffer, L. Guttman, E. A. Suchman, P.F. Lazarsfeld, S.A. Star, y J.A. Clausen (Eds.), *Measurement and prediction* (pp. 60-90). Princeton, NJ: Princeton University Press.
- [98] Haberman, S. J., von Davier, M., & Lee, Y.-H. (2008). Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions (Research Report No. RR-08-45). Princeton, NJ: Educational Testing Service.
- [99] Haberman, S. J., & Sinharay, S. (2010a). How can multivariate item response theory be used in reporting of subscores (Research Report No. 42 RR-10-09). Princeton, NJ: Educational Testing Service.
- [100] Haberman, S. J., & Sinharay, S. (2010b). Reporting of subscores using multidimensional item response theory. *Psychometrika*, 75, 209-227.
- [101] Hahne, J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*, 50, 379-390.
- [102] Haley, D.C. (1952): *Estimation of the dosage mortality relationship when the dose is subject to error*, (Technical Report N° 15), Stanford, CA: Stanford University, Applied Mathematics and Statistics Laboratory.
- [103] Hall, P. & Huang, L.S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics*, 29, 624-647.

- [104] Hambleton, R.; Swaminathan, H. y Rogers, H. (1991): *Fundamentals of Item Response Theory*, Thousand Oaks: Sage Publications.
- [105] Hanson, D. L., Pledger, G. & Wright, F. T. (1973). On consistency in monotonic regression. *The Annals of Statistics*, 401-421.
- [106] Härdle, W., Hall, P., & Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum?. *Journal of the American Statistical Association*, 83(401), 86-95.
- [107] Härdle, W. (1990): *Applied nonparametric regression*, Cambridge: Cambridge University Press.
- [108] Härdle, W., Hall, P., & Marron, J. S. (1992). Regression smoothing parameters that are not far from their optimum. *Journal of the American Statistical Association*, 87(417), 227-233.
- [109] Härdle, W., & Müller, M. (1997). Multivariate and semiparametric kernel regression (No. 1997, 26). Discussion Papers, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.
- [110] Hemker, B. T., Sijtsma, K., Molenaar, I. W., and Junker, B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, 61, 679-693.
- [111] Hemker, B.T., Sijtsma, K., Molenaar, I. W. y Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, 62, 331-347.
- [112] Henning, H.J. (1976). Die Technik der Mokken-Skalenanalyse. *Psychologische Beiträge*, 18, 410-430.
- [113] Hensen, R. A., Templin, J. L., and Willse, J. T. 2009. Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 1919-210.
- [114] Henson, R. y Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29, 262-277.
- [115] Hessen, D. J. y Dolan, C. V. (2009). Heteroscedastic one-factor models and marginal maximum likelihood estimation. *British Journal of Mathematical and Statistical Psychology*, 62, 57-77.
- [116] Hohensinn, C., Kubinger, K. D., Reif, M., Holocher-Ertl, S., Khorramdel, L., and Frebort, M. (2008). Examining item-position effects in large-scale assessment using the linear logistic test model. *Psychology Science Quarterly*, 50(3), 391-402.
- [117] Holland, P.W. (1990): On the sampling theory foundations of item response theory models, *Psychometrika*, 55, 577-601.
- [118] Holland, P. W. y Rosenbaum, P.R. (1986): Conditional association and unidimensionality in monotone latent variable models, *The Annals of Statistics*, 14, 1523-1543.
- [119] Holland, P.W. y Thayer, D.T.(1988): Diferential item performance and the Mantel-Haenszel procedure. En H. Wainer y H.I. Braun (Eds). *Test validity*, (pp.129-145). Hillsdale, NJ: LEA.
- [120] Horn, J. L. (1965). A Rationale and Test For the Number of Factors in Factor Analysis, *Psychometrika*, 30, 179-85.
- [121] Horst, P. (1965). *Factor Analysis of Data Matrices*. New York: Holt, Rinehart and Winston.
- [122] Ishwaran, H., and James, L. 2001. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96, 161-173.
- [123] Jannarone, R.J. (1997). Models for locally dependent responses: conjunctive item response theory. En W.J. van der Linden y R. K. Hambleton (Eds.), *Handbook of modern item response theory*, pp.465-479. New York: Springer-Verlag.
- [124] Jansen, P. W. G., van den Wollenberg, A. L., and Wierda, F. W. (1988). Correcting unconditional parameter estimates in the Rasch model for inconsistency. *Applied Psychological Measurement*, 12, 297-306.
- [125] Janssen, R., Schepers, J., and Peres, D. (2004). Models with item and item group predictors. En: P. De Boeck and M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* , pp. 189-212. New York: Springer.
- [126] Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.
- [127] Johnson, M.S. (2006). Nonparametric estimation of item and respondent locations from unfolding type items. *Psychometrika*, 71 (2), 257-279.
- [128] Johnson, M.S. y Junker, B.W. (2003). Using data augmentation and Markov chain Monte Carlo for the estimation of unfolding response models. *Journal of Educational and Behavioral Statistics*, 28, 195-230.
- [129] Junker, B.W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, 56, 255-278.
- [130] Junker, B. W. (1993): Conditional association, essential independence and monotone unidimensional item response models, *The Annals of Statistics*, 21, 1359-1378.

- [131] Junker, B.W. (2001): On the interplay between nonparametric and parametric IRT, with some thoughts about the future. En A. Boomsma, M.A., J. van Duijn y T. A. B. Snijders (Eds.), *Essays on item response theory*, pp.247-276, Nueva York: Springer.
- [132] Junker, B. W. y Sijtsma, K. (2001): Nonparametric item response theory in action: An overview of the special issue, *Applied Psychological Measurement*, 25, 211-220.
- [133] Kale, B. K. (1962). On the solution of likelihood equations by iteration processes. The multiparametric case. *Biometrika*, 49, 479-486.
- [134] Kalli, M., Griffin, J., and Walker, S. 2011. Slice sampling mixture models. *Statistics and Computing*, 21, 93?105.
- [135] Karabatsos, G. 2014. Bayesian regression: Nonparametric and parametric models. Software users manual, University of Illinois-Chicago, <http://www.uic.edu/georgek/HomePage/BayesSoftware.html>.
- [136] Karabatsos, G. 2015. A menu-driven software package for Bayesian regression analysis. *ISBA Bulletin*, 22(4), 13?16.
- [137] Karabatsos, G. 2016. A menu-driven software package of Bayesian nonparametric (and parametric) mixed models for regression analysis and density estimation. *Behavior Research Methods*. Submitted for publication.
- [138] Karabatsos, G. and Walker, S. 2012a. Adaptive-modal Bayesian nonparametric regression. *Electronic Journal of Statistics*, 6, 2038?2068.
- [139] Karabatsos, G. and Walker, S. 2012b. Bayesian nonparametric mixed random utility models. *Computational Statistics and Data Analysis*, 56(6), 1714?1722.
- [140] Kelly, C. & Rice, J. (1990). Monotone smoothing with application to dose response curves and the assessment of synergism. *Biometrics*, 46, 1071-1085.
- [141] Kim, S. H. y Cohen, A. S. (1992): A comparison of two areas measures for detecting differential item functioning across manifest and latent examinee groups, *Journal of Educational Measurement*, 15(3), 269-278.
- [142] Kovacs, M. (1992). *Children's depression inventory*. North Tonawanda, NY: Multi-Health System.
- [143] Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. (*Psychology Science Quarterly*), 50, 311-327.
- [144] Kubinger, K. D. (2009). Applications of the linear logistic test model in psychometric research. (*Educational and Psychological Measurement*), 69, 232-244.
- [145] Kubinger, K. D., and Draxler, C. (2006). A comparison of the Rasch model and constrained item response theory models for pertinent psychological test data. En: M. von Davier and C. H. Carstensen (Eds.), *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*, pp. 295-312. New York: Springer.
- [146] Lazarsfeld, P. F. y Henry, N. W. (1968). *Latent Structure Analysis*. New York: Houghton Mifflin.
- [147] Lee, Y. S. (2007). A comparison of methods for nonparametric estimation of item characteristic curves for binary items. *Applied Psychological Measurement*, 31(2), 121-134.
- [148] Lee, Y. S. (2002). *Applications of isotonic regression in item response theory*. Ph. D. Dissertation. University of Wisconsin-Madison.
- [149] Leighton, J. P. y Gierl, M. J. (Eds.) (2007) *Cognitive Diagnostic Assessment in Education: Theory and Practice*. New York: Cambridge University Press.
- [150] Levine, M.V. (1984): *An introduction to multilinear formula score theory*, (Measurement Series 84-5) Champaign, Il. University of Illinois. Departament of Educational Psychology. Model-Based Measurement Laboratory.
- [151] Levine, M.V. (1985): *The trait in latent trait theory*. En D. J. Weiss (Ed) *Proceedings of the 1982 item response theory/computerized adaptative testing conference*. Minneapolis, MN: University of Minnesota. Departament of Psychology. Computerized adaptative testing Laboratory.
- [152] Levine, M. V. & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational and Behavioral Statistics*, 4(4), 269-290.
- [153] Lewis, C. (1983): Bayesian inference for latent abilities. En S. B. Anderson y J. S. Helmick (Eds.), *On Educational Testing*, pp.224-251, San Francisco: Jossey-Bass.
- [154] Ligtvoet, R., Van der Ark, L. A., Te Marvelde, J. M., and Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational and Psychological Measurement*, 70, 578?595.
- [155] Ligtvoet, R., Van der Ark, L. A., Bergsma, W. P., and Sijtsma, K. (2011). Polytomous latent scales for the investigation of the ordering of items. *Psychometrika*, 76, 200?216.

- [156] Ligtoet, R. (2012). An isotonic partial credit model for ordering subjects on the basis of their sum scores. *Psychometrika*, 77, 479-494.
- [157] Linacre, J. M. (2002). *Winsteps Rasch Measurement Computer Program*. Chicago: Winsteps.com.
- [158] Loevinger, J. (1947): A systematic approach to the construction and evaluation of tests of ability, *Psychological Monographs*, 61, N° 4.
- [159] Loevinger, J. (1948): The technique of homogeneous tests compared with some aspects of 'scale analysis' and factor analysis, *Psychological Bulletin*, 45, 507-530.
- [160] Lord, F. M. (1951). *A theory of test scores and their relation to the trait measured* (Research Bulletin No. RB-51-13). Princeton, NJ: Educational Testing Service.
- [161] Lord, F. M. (1952a). *A theory of test scores* (Psychometric Monograph No. 7). Richmond, VA: Psychometric Corporation.
- [162] Lord, F. M. (1952b). The relation of test score to the ability underlying the test (Research Bulletin No. RB-52-10). Princeton, NJ: Educational Testing Service.
- [163] Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-549.
- [164] Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 117-138.
- [165] Lord, F. M. (1980a). *Applications of item response theory to practical testing problems*, Hillsdale, NJ: Erlbaum.
- [166] Lord, F. M. (1980b). Some how and which for practical tailored testing. En L. J. Th. van der Kamp, W. F. Langerak, & D. N. M. de Gruijter, (Eds.), *Psychometrics for educational debates*, pp. 189-205. New York, NY: Wiley.
- [167] Lord, F. M. y Novick, M.R. (1968): *Statistical theories of mental tests scores*, Reading, MA: Addison-Wesley.
- [168] Luce, R. D. (1959). *Individual Choice Behavior*. New York: Wiley.
- [169] Ludlow, L. H. (1986): Graphical analysis of item response theory residuals, *Applied Psychological Measurement*, 10, 217-229.
- [170] Lumsden, J. (1961): The construction of unidimensional tests, *Psychological Bulletin*, 58, 122-131.
- [171] Luzardo, M. y Forteza, D. (2014). *Modelo no paramétrico multidimensional para la estimación de los rasgos y de las curvas características del ítem mediante regresión no paramétrica con núcleos*. Montevideo: CSIC.
- [172] Luzardo, M., y Rodríguez, P. (2015). A Nonparametric Estimator of a Monotone Item Characteristic Curve. En L. A. van der Ark, D. Bolt, W.Ch. Wang, J. A. Douglas y S. M. Chow (Eds.), *Quantitative Psychology Research*, pp. 99-108. Suiza: Springer International Publishing.
- [173] MacEachern, S. 1999. Dependent nonparametric processes. Proceedings of the Bayesian Statistical Sciences Section of the American Statistical Association, Alexandria, VA, 50?55.
- [174] MacEachern, S. 2000. Dependent Dirichlet processes, Technical Report, Department of Statistics, Columbus, OH: The Ohio State University.
- [175] MacEachern, S. 2001. Decision theoretic aspects of dependent nonparametric processes. In E. George (Ed.), *Bayesian Methods with Applications to Science, Policy and Official Statistics* (pp. 551?560). Creta, Greece: International Society for Bayesian Analysis.
- [176] Mair, P. y Hatzinger, R. (2007). Extended Rasch modelling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20, 1-20.
- [177] Mammen, E. (1991). Estimating a smooth monotone regression function. *Annals of Statistics*, 19, 724-740.
- [178] Marconi, C., Méndez, I., Silva, H. y Brovetto, C. (2015). *Evaluación Adaptativa de aprendizajes de inglés en Educación Primaria de Uruguay, 2015*. Montevideo: Plan Ceibal.
- [179] Massart, P. (2000): About the constants in Talagrand's concentration inequalities for empirical processes, *Annals of Probability*, 28, 863-884.
- [180] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- [181] Masters, G. N. (1987). Measurement models for ordered response categories. En Langeheine, R., y Rost, J. (Eds.), *Latent Trait and Latent Class Models*, pp. 11-29. New York: Plenum Publishing Corporation.
- [182] Masters, G. N. (1988a). Partial credit model. In Keeves, J.P. (Ed.) *Educational Research Methodology, Measurement, and Evaluation*, pp. 292-296. Oxford: Pergamon Press.

- [183] Masters, G. N. (1988b). The analysis of partial credit scoring. *Applied Measurement in Education*, 1, 279-298.
- [184] Masters, G. N. (2010). Partial credit model. En Nering, M., and Ostini, R. (Eds.), *Handbook of Polytomous Item Response Theory Models: Development and Applications*, pp. 109-122. New York: Routledge Academic.
- [185] Masters, G. N. (2016). Partial Credit Model. En: W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume one: Models (pp.109-119). Boca Raton, FL: Chapman & Hall/CRC Press.
- [186] Masters, G. N., and Wright, B. D. (1984). The essential process in a family of measurement models. *Psychometrika*, 49, 529-544.
- [187] Masters, G. N., and Wright, B. D. (1996). The partial credit model. In Hambleton, R. K., and van der Linden, W. J. (Eds.), *Handbook of Modern Item Response Theory*, pp. 101-119. New York: Springer.
- [188] Maxwell, A. E. (1959): Maximum likelihood estimates of item parameters using the logistic function, *Psychometrika*, 24, 221-227.
- [189] Mazor, K. M., Clauser, B. E. y Hambleton, R. K. (1994): Identification for nonuniform differential item functioning using a variation of the Mantel- Haenszel procedure, *Educational and Psychological Measurement*, 54(2), 284-291.
- [190] McDiarmid, C. (1989): *On the method of bounded differences*. En: *Surveys in Combinatorics 1989*. Cambridge University Press, Cambridge, 148-188.
- [191] McDonald, R. P. (1962a). A note on the derivation of the general latent class model. *Psychometrika*, 27, 203-206.
- [192] McDonald, R. P. (1967). *Nonlinear factor analysis*. Psychometric Monographs, No. 15.
- [193] McDonald, R. P. (1981): The dimensionality of tests and items, *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- [194] McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement*, 4, 379-396.
- [195] McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- [196] McKinley, R.L y Mills, C.N. (1985): A comparison of several goodness-of-fit statistics. *Applied Psychological Measurement* 9 (1), 49-57.
- [197] Meijer, R.R., Sijtsma, K. y Smid, N.G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 11, 283-298.
- [198] Mellenbergh, G. J. (2016) Models for Continuous Responses. En: W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume one: Models, pp.153-163. Boca Raton, FL: Chapman & Hall/CRC Press.
- [199] Millsap, R. E. y Everson, H. T. (1993): Methodology review: statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 16, 389-402.
- [200] Mislevy, R.J. (1986a): Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, 11, 3-31.
- [201] Mislevy, R.J. (1986b): Bayes modal estimation in item response models *Psychometrika*, 51, 177-195.
- [202] Mislevy, R. J. (1988). Exploiting auxiliary information about items in the estimation of Rasch item difficulty parameters. *Applied Psychological Measurement*, 12, 725-737.
- [203] Mislevy, R.J. y Bock, R.D. (1982): *Bilog Item analysis and test scoring with binary logistic models*. Mooresville, ID: Scientific Software.
- [204] Mokken, R. J. (1971): *A theory and procedure of scale analysis*, The Hague, Mouton/Berlin: De Gruyter.
- [205] Mokken, R.J. (1997): Nonparametric models for dichotomous responses. En W. J. van der Linden y R. K. Hambleton (eds.) *Handbook of modern item response theory*, pp. 351-367, Nueva York: Springer.
- [206] Mokken, R. J. y Lewis, C. (1982): A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417-430.
- [207] Molenaar, I.W. (1997): Nonparametric models for polytomous responses. En W. J. van der Linden y R. K. Hambleton (Eds.) *Handbook of modern item response theory*, pp. 369-380, Nueva York: Springer.
- [208] Molenaar, D., Dolan, C. V., y Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, 63, 293-317.
- [209] Molenaar, I.W. y Sijtsma, K. (2000): *User's manual MSP5 for Windows*, Groningen, iecProGAMMA.

- [210] Mukerjee, H.(1988). Monotone nonparametric regression. *Annals of Statistics*, 16, 741-750.
- [211] Mulaik, S. A. (1972). *A mathematical investigation of some multidimensional Rasch models for psychological tests*. Ponencia presentada en el Meeting of the Psychometric Society, Princeton, NJ.
- [212] Müller, H. (1987). A Rasch model for continuous ratings. *Psychometrika*, 52, 165-181.
- [213] Muñoz, J. (1997): *Introducción a la teoría de respuesta a los ítems*, Madrid, Pirámide.
- [214] Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- [215] Muraki, E. y Muraki, M. (2016). Generalized Partial Credit Model. En: W. J. van der Linden (Ed.). (2016). *Handbook of Item Response Theory*, Volume one: Models, pp.127-135. Boca Raton, FL: Chapman & Hall/CRC Press.
- [216] Muraki, E. and Wang, M. (1992). Issues relating to the marginal maximum likelihood estimation of the partial credit model. *Paper presented at the Annual Meeting of the American Educational Research Association*, San Francisco, California.
- [217] Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.
- [218] Nadaraya, E. A. (1964): On estimating regression. *Theory of Probability and its Applications* 10, 186-190.
- [219] Niemöller, B y van Schuur, W. H. (1983). Stochastic models for unidimensional scaling: Mokken and Rasch. En D. Mc Kay, N. Schofield y P. Whiteley (Eds.), *Data Analysis and the Social Sciences*, pp.120-170. London: Francis Pinter.
- [220] Neyman, J y Scott, E.L. (1948): Consistent estimates based on partially consistent observations. *Econometrika* 16 (1), 1-32.
- [221] Park, D. G. (1988): *Investigations of item response theory item bias detection*, Universidad de Georgia. Tesis doctoral no publicada.
- [222] Parzen, E. (1962): On estimation of a probability density and mode, *Annals of Mathematical Statistics*, 35, 1065-1076.
- [223] Post, W.J. (1992). *Nonparametric unfolding models: A latent structure approach*. Leiden, Holanda: DSWO Press.
- [224] Raju, N. S. (1990): Determining the significance of estimated signed and unsigned areas between two item response functions, *Applied Psychological Measurement*, 14(2), 197-207.
- [225] Ramsay, J. O.(1991): Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika*, 56, 611-630.
- [226] Ramsay, J.O. (1994): *TESTGRAF: A program for the graphical analysis of multiple choice test and questionnaire data*. User´s guide.
- [227] Ramsay, J.O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 60, 365-375.
- [228] Ramsay, J. O., y Abrahamowicz, M. (1989): Binomial regression with monotone splines. A psychometric application. *Journal of the American Statistical Association*, 84, 906-915.
- [229] Ramsay, J. O. y Winsberg, S. (1991): Maximum marginal likelihood estimation for semiparametric item analysis, *Psychometrika*, 56, 365-379.
- [230] Ramsay, J. O. 1996. A geometrical approach to item response theory. *Behaviormetrika*, 23, 3?17.
- [231] Ramsay, J. O. and Silverman, B. W. 2002. *Applied Functional Data Analysis*. New York: Springer.
- [232] Ramsay, J. O. and Silverman, B. W. 2005. *Functional Data Analysis*. New York: Springer.
- [233] Ramsay, J. O., Hooker, G., and Graves, S. 2009. *Functional Data Analysis with R and MATLAB*. New York: Springer.
- [234] Ramsay, J.O. (2016) Functional Approaches to Modeling Response Data. En: W. J. van der Linden (Ed.). *Handbook of Item Response Theory*, Volume one: Models, pp. 337-352. Boca Raton, FL: Chapman & Hall/CRC Press.
- [235] Ramsay, J. O. and Wiberg, M. 2016. Nonparametric Estimation of Item Characteristic Curves Using Parameter Cascading. Unpublished manuscript.
- [236] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.
- [237] Rasch, G. (1961). On general laws and meaning of measurement in psychology. En: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 321-333.
- [238] aykov, T. and Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. New York: Routledge.
- [239] Reckase, M. D. (2009). *Multidimensional Item Response Theory*. New York: Springer.

- [240] Reise, S. P. y Waller, N.G.(1990): Fitting the two-parameter model to personality data, *Applied Psychological Measurement*, 14, 45-58.
- [241] Revelle, W. & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14(4), 403-414.
- [242] Richardson, M. W. (1936): The relationship between difficulty and the differential validity of a test, *Psychometrika* 1, 33-49.
- [243] Rijmen, F. y De Boeck, P. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, 26, 271-285.
- [244] Rijmen, F. (2009b). Three multidimensional models for testlet-based tests: Formal relations and an empirical comparison (Research Report No. RR-09-37). Princeton, NJ: Educational Testing Service.
- [245] Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47, 361-372.
- [246] Rijmen, F., Jeon, M., von Davier, M., & Rabe-Hesketh, S. (2013). A general psychometric approach for educational survey assessments: Flexible statistical models and efficient estimation methods. En L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *A handbook of international large-scale assessment data analysis*. London, England: Chapman & Hall.
- [247] Roberts, J.S., Donoghue, J.R. y Laughlin, J.E. (2000). A general model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement*, 24, 3-32.
- [248] Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York: Wiley.
- [249] Rodríguez, P. (2016) *Creación y establecimiento de estándares para la evaluación de la calidad de la educación superior: un modelo adaptado a los centros universitarios de la Udelar*. Tesis doctoral. Universidad Nacional de Educación a Distancia, Madrid, España. Recuperado de <http://e-spacio.uned.es/fez/view/tesisuned:Educacion-Pcrodriguez>.
- [250] Rodríguez, P. (2017). Creación, desarrollo y resultados de la aplicación de pruebas de evaluación basadas en estándares para diagnosticar competencias en Matemática y Lectura al ingreso a la Universidad. *Revista Iberoamericana de Evaluación Educativa*, 10 (1), 89-107. <https://doi.org/10.15366/rie2017.10.1.005>.
- [251] Rose, N., von Davier, M., & Xu, X. (2010). Modeling nonignorable missing data with item response theory (IRT) (Research Report No. RR-10-11). Princeton, NJ: Educational Testing Service.
- [252] Rosenbaum, P. (1984): Testing the conditional independence and monotonicity assumptions of item response theory, *Psychometrika*, 49, 425-435.
- [253] Rosenbaum, P. (1987): Comparing item characteristic curves, *Psychometrika*, 52(2), 217-233.
- [254] Rosenblatt, M. (1956): Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics*, 27, 642-669.
- [255] Rossi, N., Wang, X., and Ramsay, J. O. 2002. Nonparametric item response function estimates with the EM algorithm. *Journal of the Educational and Behavioral Sciences*, 27, 291-317.
- [256] Rost, J. 1990. Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- [257] Rost, J. 1991. A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology*, 44, 75-92.
- [258] Rost, J. y von Davier, M. (1994): A conditional item-fit index for Rasch models, *Applied Psychological Measurement*, 18, 171-182.
- [259] Rupp, A. A., Templin, J., and Henson, R. A. (2010). *Diagnostic Measurement*. New York: Guilford Press.
- [260] Samejima, F. (1969). Estimation of ability using a response pattern of graded scores. *Psychometric Monograph No. 17*. Richmond, VA: Psychometric Corporation. Recuperado de: <http://www.psychometrika.org/journal/online/MN07.pdf>
- [261] Samejima, F. (1972). A general model for free-response data. *Psychometric Monograph No. 18*. Richmond, VA: Psychometric Corporation. Recuperado de: <http://www.psychometrika.org/journal/online/MN07.pdf>
- [262] Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, 38, 203-219.
- [263] Samejima, F. (1977). A method of estimating item characteristic functions using the maximum likelihood estimate of ability. *Psychometrika*, 42, 163-191.
- [264] Samejima, F. (1993). The bias function of maximum likelihood estimate of ability for the dichotomous response level. *Psychometrika*, 58, 195-209.

- [265] Samejima, F. (1995). Acceleration model in the heterogeneous case of the general graded response model. *Psychometrika*, 60, 549-572.
- [266] Samejima, F. (1997). Graded response model. In W. J. van der Linden and R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, pp. 85-100. New York: Springer.
- [267] Samejima, F. (1998). Efficient nonparametric approaches for estimating the operating characteristics of discrete item responses. *Psychometrika*, 63, 111-130.
- [268] Samejima, F. (2000). Logistic positive exponent family of models: Virtue of asymmetric item characteristic curves. *Psychometrika*, 65, 319-335.
- [269] Samejima, F. (2004). Graded response model. En K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement*. Volume 2, pp. 145-153. Amsterdam: Elsevier.
- [270] Samejima, F. (2016). Graded Response Models. En: W. J. van der Linden (Ed.). (2016). *Handbook of Item Response Theory*, Volume one: Models. pp.95-108. Boca Raton, FL: Chapman & Hall/CRC Press.
- [271] San Martín, E. (2016). Identification of Item Response Theory Models. En: W. J. van der Linden (Ed.). (2016). *Handbook of Item Response Theory*, Volume two: Statistical Tools. pp.127-150. Boca Raton, FL: Chapman & Hall/CRC Press.
- [272] Santisteban, C. y Alvarado, J. (2001): *Modelos Psicométricos*. Madrid, UNED.
- [273] Schriever, B. F. (1985): *Order Dependence*. Tesis doctoral no publicada. Free University, Amsterdam.
- [274] Segall, D.O. (1983): *Test characteristic curves item bias, and transformation to a common metric in item response theory: A methodological artifact with serious consequences and a simple solution*. Urbana-Champaign: University of Illinois, Department of Psychology. Manuscrito no publicado.
- [275] Serfling, R. J. (1980): *Approximations Theorems of mathematical statistics*, New York: Wiley.
- [276] Sethuraman, J. 1994. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639-650.
- [277] Shealy, R. y Stout, W. (1993a): A model-based standarization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- [278] Shealy, R. y Stout, W. (1993b): A model-based standarization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF, *Psychometrika*, 58, 159-194.
- [279] Shealy, R., Stout, W. y Roussos, L. (1995): SIBTEST (versión 1.3), Urbana-Champaign, University of Illinois, Department of Applied Statistics.
- [280] Silverman, B. W. (1986). Density estimation for statistics and data analysis (Vol. 26). CRC press.
- [281] Sinharay, S., Johnson, M. S., & Williamson, D. (2003). An application of a Bayesian hierarchical model for item family calibration (Research Report No. RR-03-04). Princeton, NJ: Educational Testing Service.
- [282] Sijtsma, K. (1988): *Contributions to Mokken's nonparametric item response theory*. Amsterdam: Free University Press.
- [283] Sijtsma, K. (1998): Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores, *Applied Psychological Measurement*, 22, 3-32.
- [284] Sijtsma, K. (2001): Developments in measurement of persons and items by means of item response models, *Behaviormetrika*, 28, 65-94.
- [285] Sijtsma, K. y Molenaar, I. W. (2002): *Introduction to nonparametric item response theory*. Thousand Oaks: Sage Publications.
- [286] Sijtsma, K. y Junker, B.W. (2006): Item Response Theory: past performance, present developments, and future expectations. *Behaviormetrika*, 33 (1), 75-102.
- [287] Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and Reliability of Situational Judgement Test Scores A New Approach Based on Cognitive Diagnosis Models. *Organizational Research Methods*, 19(3), 506-532.
- [288] Sperlich, S. (2009). A note on nonparametric estimation with predicted variables. *The Econometrics Journal*, 12(2), 382-395.
- [289] Spray, J. A. y Carlson, J. E. (1986): *Comparison of loglinear and logistic regression models for detecting changes in proportions*. Comunicación presentada a la Annual Meeting of the American Educational Research Association, San Francisco.
- [290] Stout, W. F. (1987): A nonparametric approach for assessing latent trait unidimensionality, *Psychometrika*, 52, 589-617.
- [291] Stout, W. F. (1990): A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation, *Psychometrika*, 55, 293-325.

- [292] Stout, W. F. 2002. Psychometrics: From practice to theory and back. *Psychometrika*, 67, 485-518.
- [293] Straat, J. H., Van der Ark, L. A., and Sijtsma, K. 2016. Using conditional association to identify locally independent item sets. Submitted for publication.
- [294] Swaminathan, H. y Gifford, J. A. (1986): Bayesian estimation in the three-parameter logistic model, *Psychometrika*, 51, 589-601.
- [295] Takane, Y. y de Leeuw, J. (1987): On the relationship between item response theory and factor analysis of discretized variables, *Psychometrika* 52, 393-408.
- [296] Thissen, D. y Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567-577.
- [297] Thissen, D. y Steinberg, L. (2010). Using item response theory to disentangle constructs at different levels of generality. En: S. Embretson (Ed.), *Measuring Psychological Constructs: Advances in Model-Based Approaches*, pp. 123-144. Washington, DC: American Psychological Association.
- [298] Thissen, D. y Cai, L. (2016). Nominal categories models. En: W. J. van der Linden (Ed.), *Handbook of Item Response Theory*, Volume one: Models, pp.51-74. Boca Raton, FL: Chapman & Hall/CRC Press.
- [299] Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273-286.
- [300] Thurstone, L. L. and Chave, E. J. (1929). *The Measurement of Attitude*. Chicago: University of Chicago Press.
- [301] Tijmstra, J., Hessen, D. J., Van der Heijden, P. G. M., and Sijtsma, K. (2011). Invariant ordering of itemtotal regressions. *Psychometrika*, 76, 217-227.
- [302] Tucker, L. R. (1946): Maximum validity of a test with equivalent items, *Psychometrika* 11, 1-13.
- [303] Van Abswoude, A. A. H., Van der Ark, L. A., and Sijtsma, K. 2004. A comparative study of test dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement*, 28, 3-24.
- [304] Van der Ark, L.A. (2005). Practical consequences of stochastic ordering of the latent trait under various polytomous IRT models. *Psychometrika*, 70, 283-304.
- [305] Van der Ark, L. A., and Bergsma, W. P. (2010). A note on stochastic ordering of the latent trait using the sum of polytomous item scores. *Psychometrika*, 75, 272-279.
- [306] Van der Ark, L. A., Croon, M. A., and Sijtsma, K. (2008). Mokken scale analysis for dichotomous items using marginal models. *Psychometrika*, 73, 183-208.
- [307] Van der Linden, W.J. (2005). *Linear models for optimal test design*. New York: Springer.
- [308] Van der Linden, W. y Hambleton, R. (1997): *Handbook of Modern Item Response Theory*, New York: Springer-Verlag.
- [309] Van Rijn, P., y Rijmen, F. (2012). A note on explaining away and paradoxical results in multidimensional item response theory (Research Report No. RR-12-13). Princeton, NJ: Educational Testing Service.
- [310] Velicer, W.F. (1976) Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327. doi:10.1007/BF02293557
- [311] Verhelst, N.D., Glas, C.A.W. y van der Sluis, A. (1984): Estimation problems in the Rasch model>the basic symmetric functions, *Computational Statistics Quarterly*, 1, 245-262.
- [312] Verhelst, N.D. y Molenaar, W. (1988): Logit-based parameter estimation in Rasch model. *Statistica Neerlandica*, 42, 273-295.
- [313] Verhelst, N.D. y Verstralen, H.H.F.M. (1993). A stochastic unfolding model derived from the partial credit model. *Kwantitative Methoden*, 42, 195-215.
- [314] Verhelst, N. D., Glas, C., and de Vries, H. 1997. A steps model to analyse partial credit. In W. J. van der Linden and R. Hambleton (Eds.), *Handbook of Modern Item Response Theory*. New York: Springer.
- [315] von Davier, M. (2005). A general diagnostic model applied to language testing data (Research Report No. RR-05-16). Princeton, NJ: Educational Testing Service.
- [316] von Davier, M. (2007a). *Hierarchical general diagnostic models (Research Report No. RR-07-19)*. Princeton, NJ: Educational Testing Service.
- [317] von Davier, M. (2007b). *Mixture distribution diagnostic models (Research Report No. RR-07-32)*. Princeton, NJ: Educational Testing Service.
- [318] von Davier, M. (2008a). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287-307.

- [319] von Davier, M. (2008b). The mixture general diagnostic model. En G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*, pp. 255-274. Charlotte, NC: Information Age Publishing.
- [320] von Davier, M., & von Davier, A. A. (2004). *A unified approach to IRT scale linking and scale transformation (Research Report No. RR-04-09)*. Princeton, NJ: Educational Testing Service.
- [321] von Davier, M., & von Davier, A. A. (2007). A unified approach to IRT scale linking and scale transformation. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 3, 115-124.
- [322] von Davier, M., & von Davier, A. A. (2010). A general model for IRT scale linking and scale transformation. En A. A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking*, pp. 225-242. New York, NY: Springer.
- [323] Wahba, G (1990): *Spline Models for observational data*. Philadelphia. Society for Industrial and Applied Mathematics.
- [324] Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. Crc Press.
- [325] Wang, X., Bradlow, E. T., & Wainer, H. (2002). *A general Bayesian model for testlets: Theory and application (Research Report No. RR-02-02)*. Princeton, NJ: Educational Testing Service.
- [326] Watson, G. S. (1964): *Smooth regression analysis Sankhya*, Series A, 26, 101-116.
- [327] Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.
- [328] Wingersky, M.S. Patrick, R. y Lord, F.M. (1988): *Logist users guide*. Princeton, NJ: Educational Testing Service.
- [329] Wollenberg, A. L. van den (1982): A simple and effective method to test the dimensionality axiom of the Rasch model, *Applied Psychological Measurement*, 6, 83-91.
- [330] Wollenberg, A. L. van den (1988): Testing a latent trait model. En R. Langeheine y J. Rost (Ed.), *Latent trait and latent class models*, pp. 31-50. New York: Plenum Press.
- [331] Wood, R. L., Wingersky, M.S. y Lord, F. M. (1976): *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (RM-76-6)*, Princeton, NJ: Educational Testing Service.
- [332] Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *The Annals of Statistics*, 9(2), 443-448.
- [333] Wright, B. D. (1988). The efficacy of unconditional maximum likelihood bias correction: Comment on Jansen, van den Wollenberg, and Wierda. *Applied Psychological Measurement*, 12, 315-324.
- [334] Wright, B. D. y Mead, R. J. (1978): *BICAL: Calibrating items with the Rasch model* (Research Memorandum No. 23A), Chicago, IL: University of Chicago, Statistical Laboratory.
- [335] Wright, B.D. y Douglas, G. A. (1977a): Best procedures for sample-free item analysis. *Applied Psychological Measurement*, 1, 281-295.
- [336] Wright, B.D. y Douglas, G. A. (1977b): Conditional versus unconditional procedures for sample-free analysis. *Educational and Psychological Measurement*, 37, 573-586.
- [337] Wright, B. D. y Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.
- [338] Wright, B.D. y Panchapakesan, N. (1969): A procedure for sample-free item analysis, *Educational and Psychological Measurement*, 29, 23-48.
- [339] Wright, B. D. y Stone, M. H. (1979): *Best test design*, Chicago: MESA Press University of Chicago.
- [340] Wu, M. L., Adams, R. J., Wilson, M. R., and Haldane, S. A. (2007). *ACERConQuest Version 2: Generalized Item Response Modelling Software* [Computer program]. Camberwell: Australian Council for Educational Research.
- [341] Xie, Y., y Wilson, M. (2008). Investigating DIF and extensions using an LLTM approach and also an individual differences approach: an international testing context. *Psychology Science Quarterly*, 50, 403-416.
- [342] Xu, X. (2007). *Monotone properties of a general diagnostic model* (Research Report No. RR-07-25). Princeton, NJ: Educational Testing Service.
- [343] Xu, X., & von Davier, M. (2008c). Linking for the general diagnostic model (Research Report No. RR-08-08). Princeton, NJ: Educational Testing Service.
- [344] Xu, X., & von Davier, M. (2008d). Linking for the general diagnostic model. IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments 1, 97-111.
- [345] Xu, X., & von Davier, M. (2008a). Comparing multiple-group multinomial log-linear models for multidimensional skill distributions in the general diagnostic model (Research Report No. RR-08-35). Princeton, NJ: Educational Testing Service.

- [346] Yan, D., Almond, R. G., & Mislevy, R. J. (2004). *A comparison of two models for cognitive diagnosis* (Research Report No. RR-04-02). Princeton, NJ: Educational Testing Service.
- [347] Yen, M. W. (1981): Using simulation results to choose a latent trait model, *Applied Psychological Measurement*, 5, 245-262.
- [348] Zhang, J. y Stout, W. F. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64, 129-152.
- [349] Zhang, J. (2004a). Comparison of unidimensional and multidimensional approaches to IRT parameter estimation (Research Report No. RR-04-44). Princeton, NJ: Educational Testing Service.
- [350] Zhang, J. (2004b). Conditional covariance theory and DETECT for polytomous items (Research Report No. RR-04-50). Princeton, NJ: Educational Testing Service.
- [351] Zhang, J. (2005b). Estimating multidimensional item response models with mixed structure (Research Report No. RR-05-04). Princeton, NJ: Educational Testing Service.
- [352] Zhang, J. (2007). Conditional covariance theory and DETECT for polytomous items. *Psychometrika*, 72, 69-91.

Capítulo 8

Anexo I Gráficas de las Aplicaciones

8.1. CCI del cuadernillo 1 de matemática

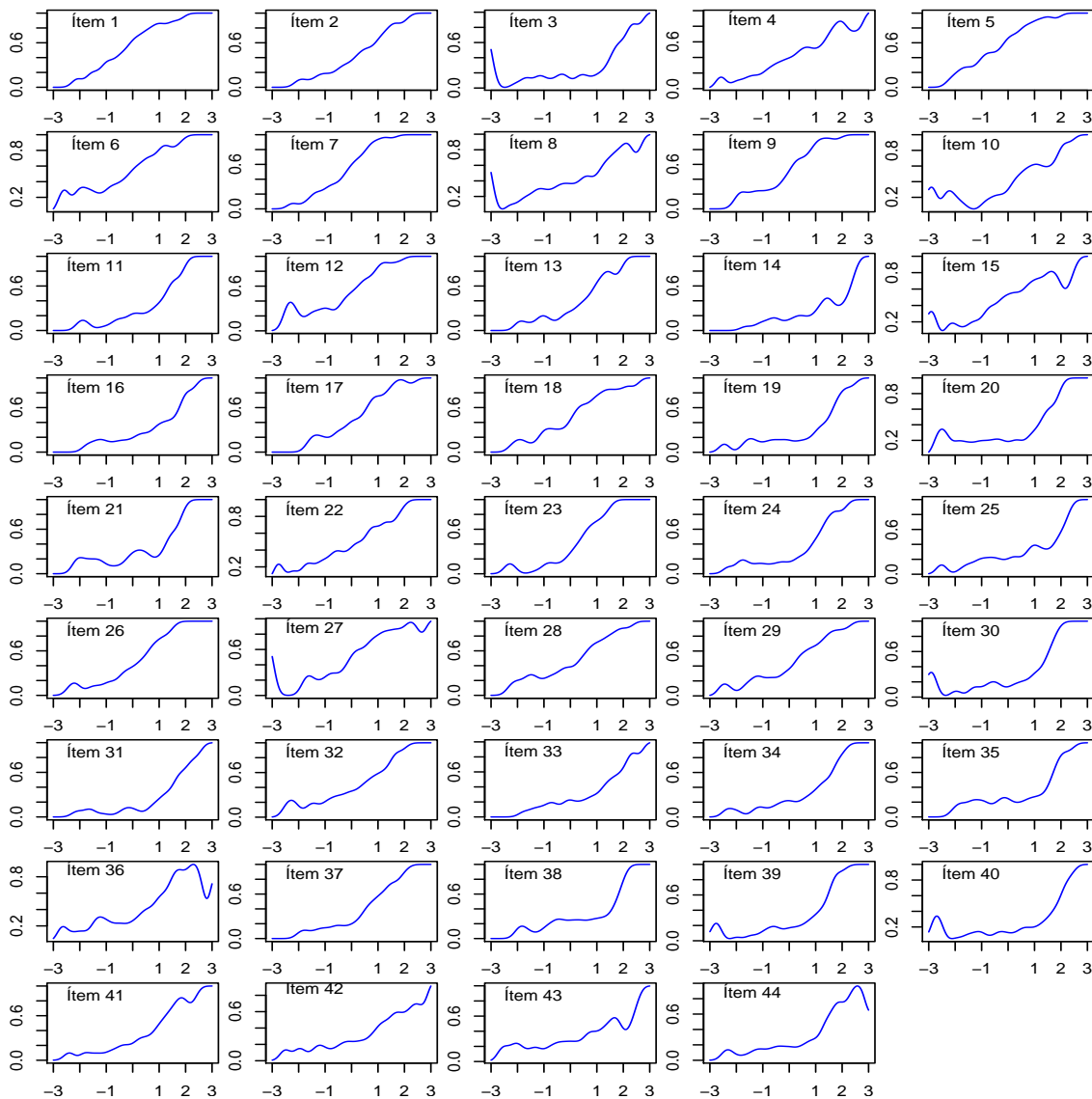


Figura 8.1: Modelo no paramétrico

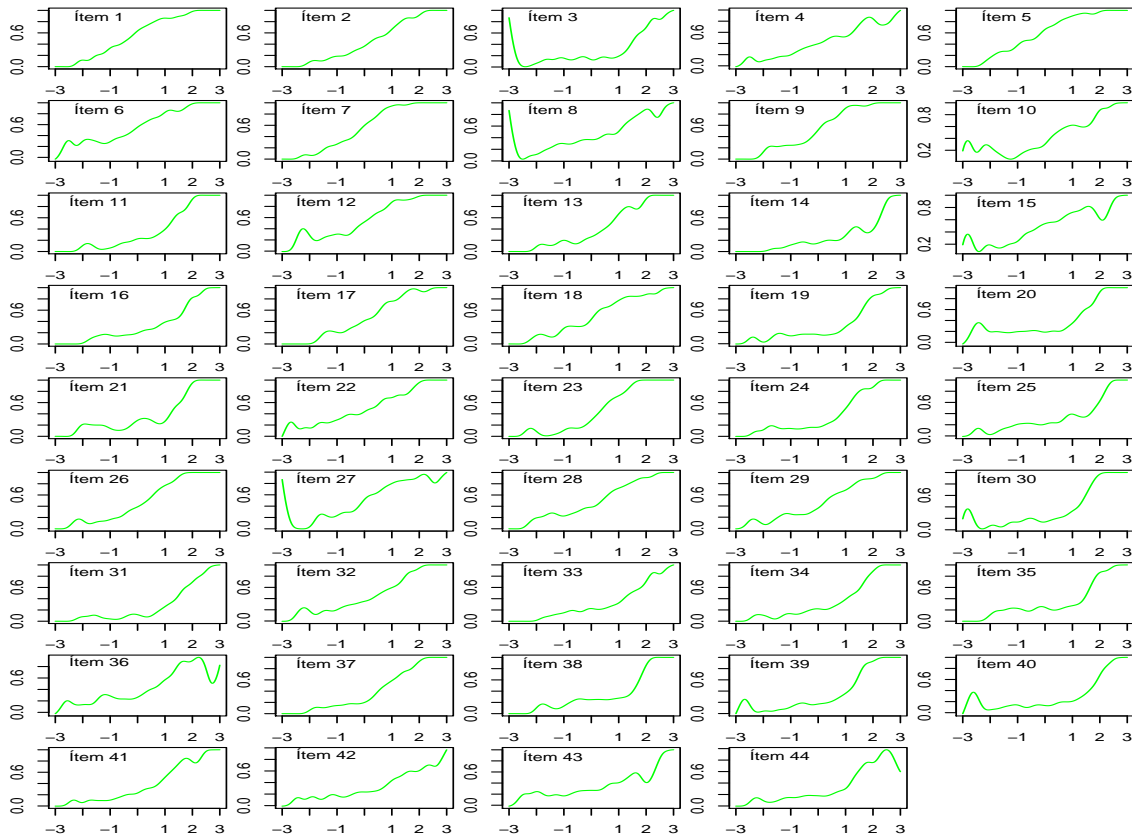


Figura 8.2: Modelo polinómico local

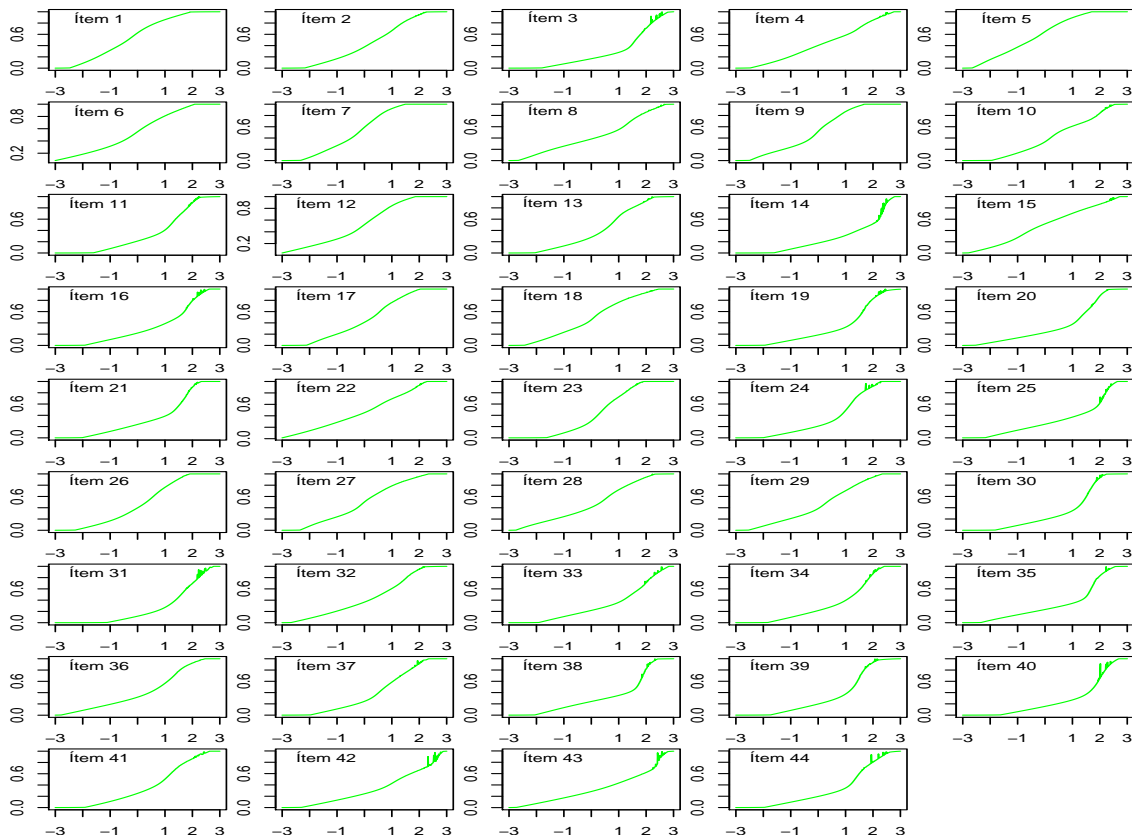


Figura 8.3: Modelo isótono usando regresión no paramétrica

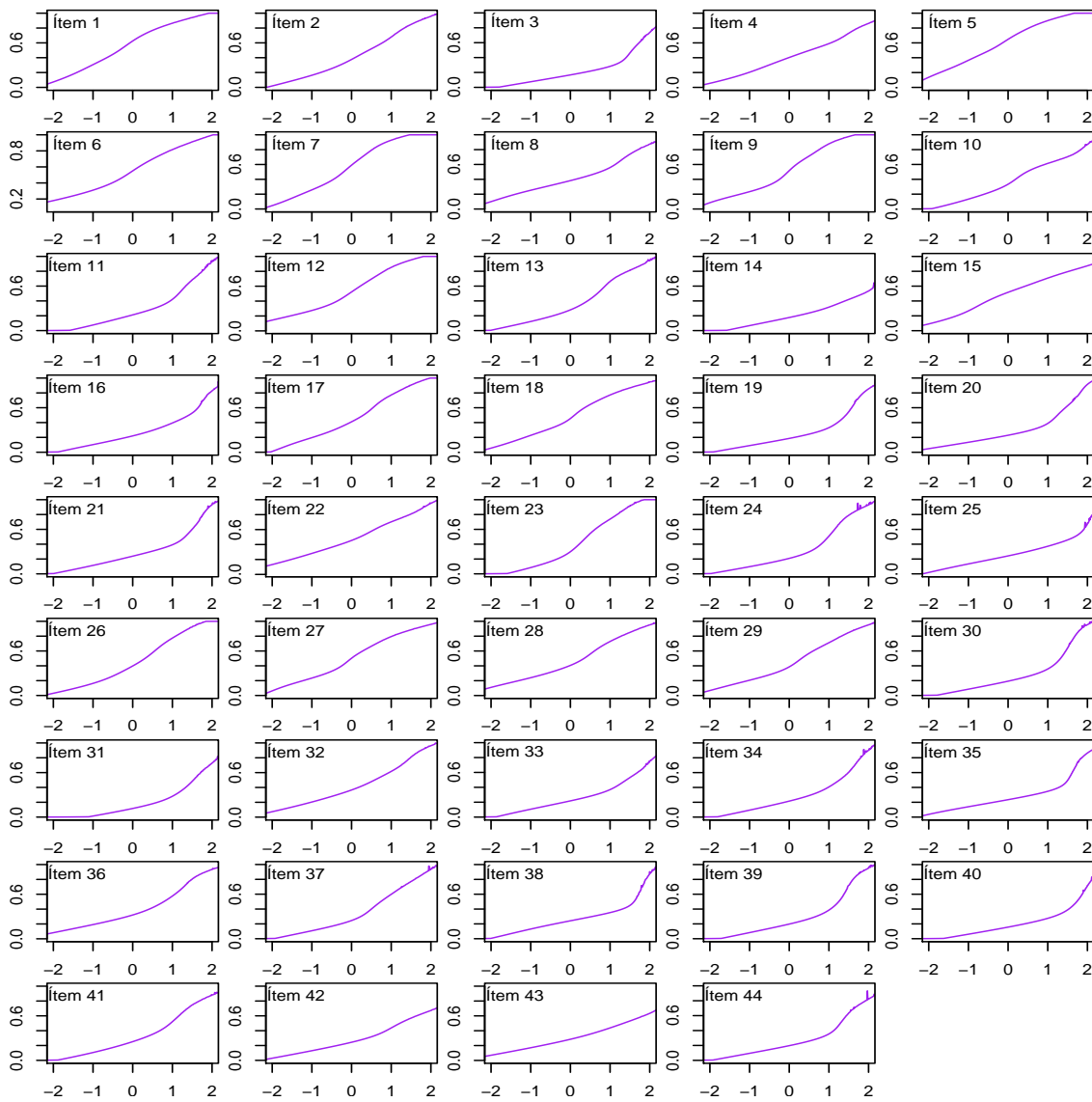


Figura 8.4: Modelo isótono usando regresión no paramétrica polinómica

8.2. Superficies estimadas en dos dimensiones

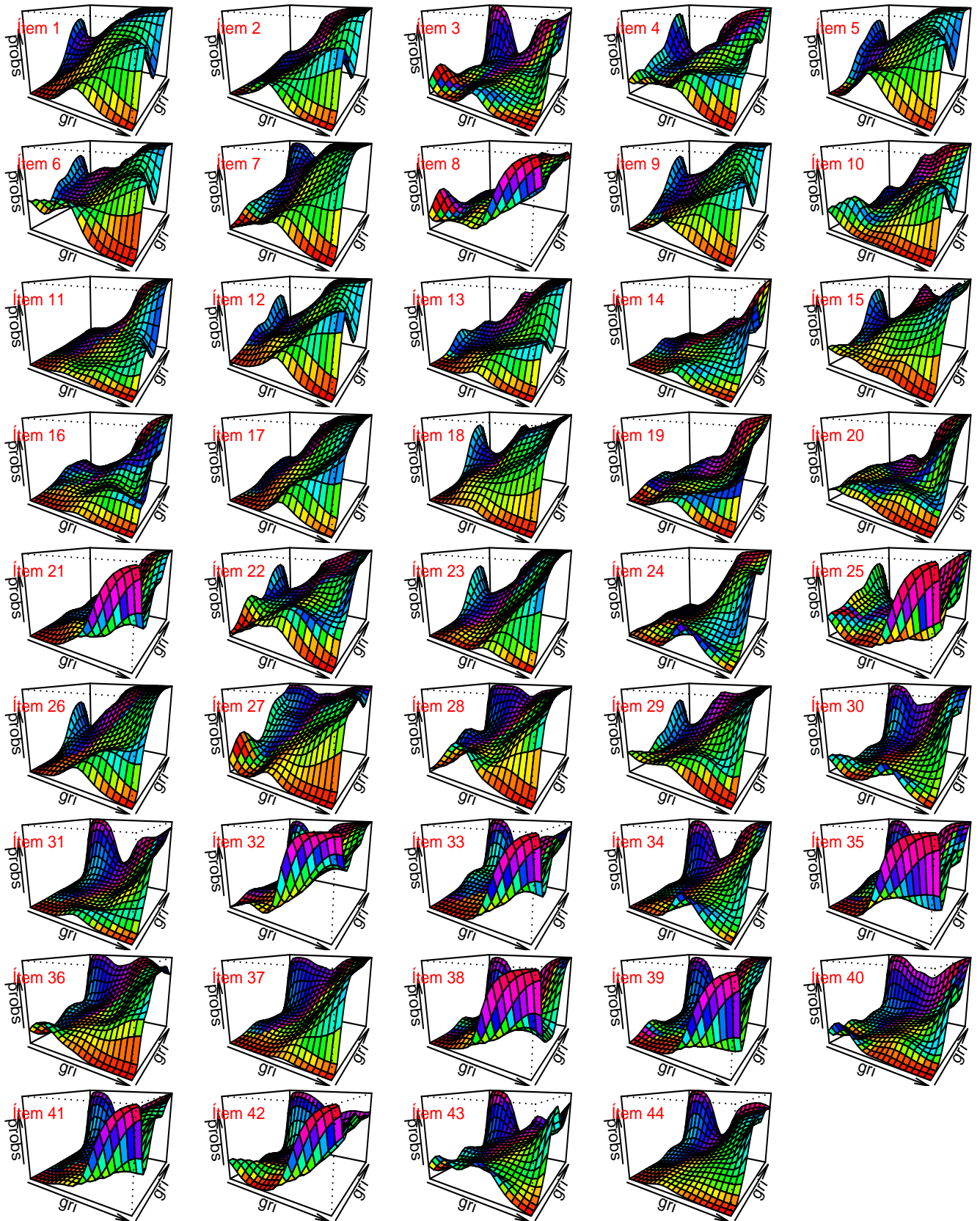


Figura 8.5: Superficies de respuesta estimadas por regresión NP

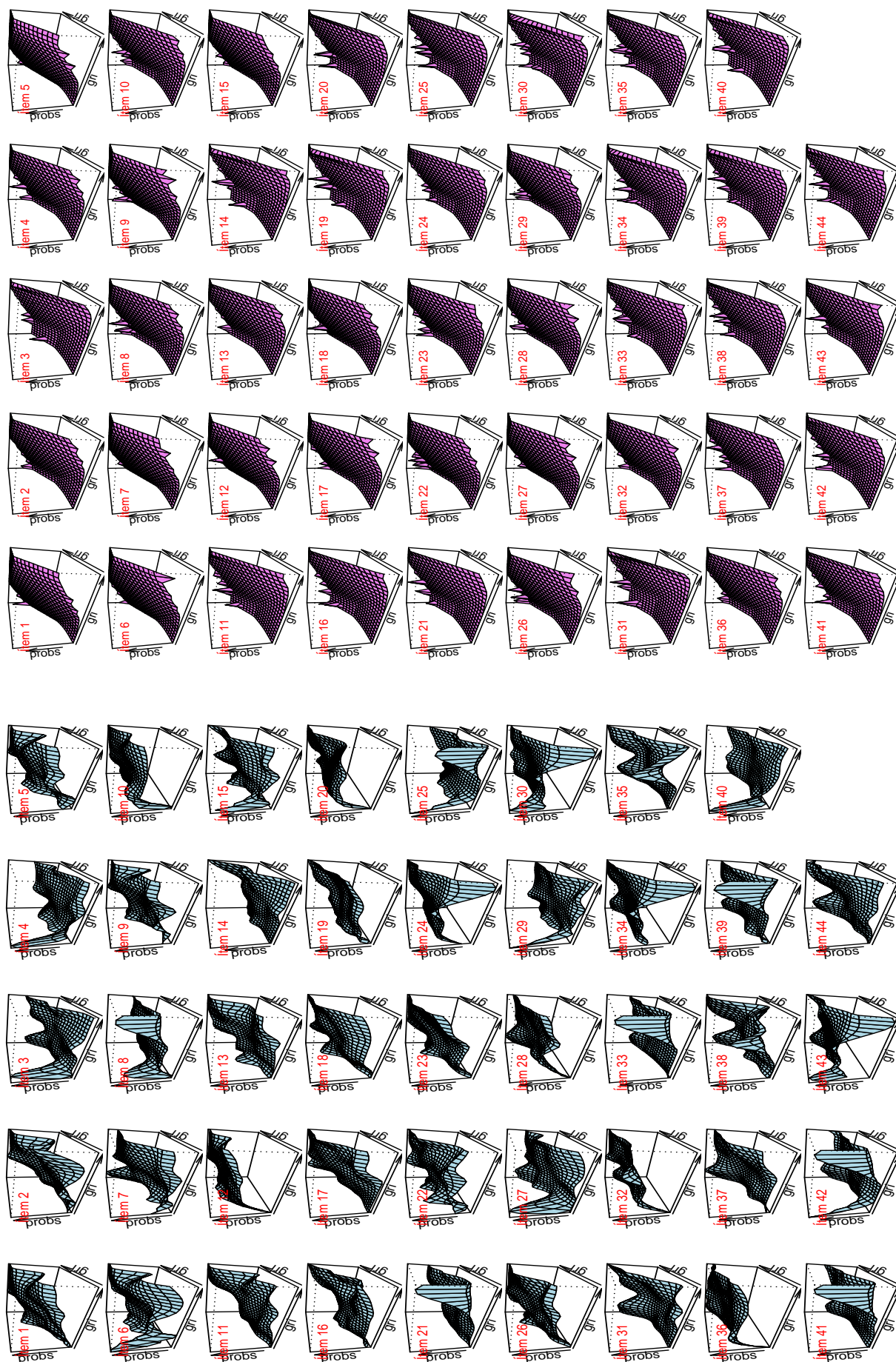


Figura 8.6: Regresión no paramétrica polinómica

Figura 8.7: Modelo isotónico usando regresión no paramétrica

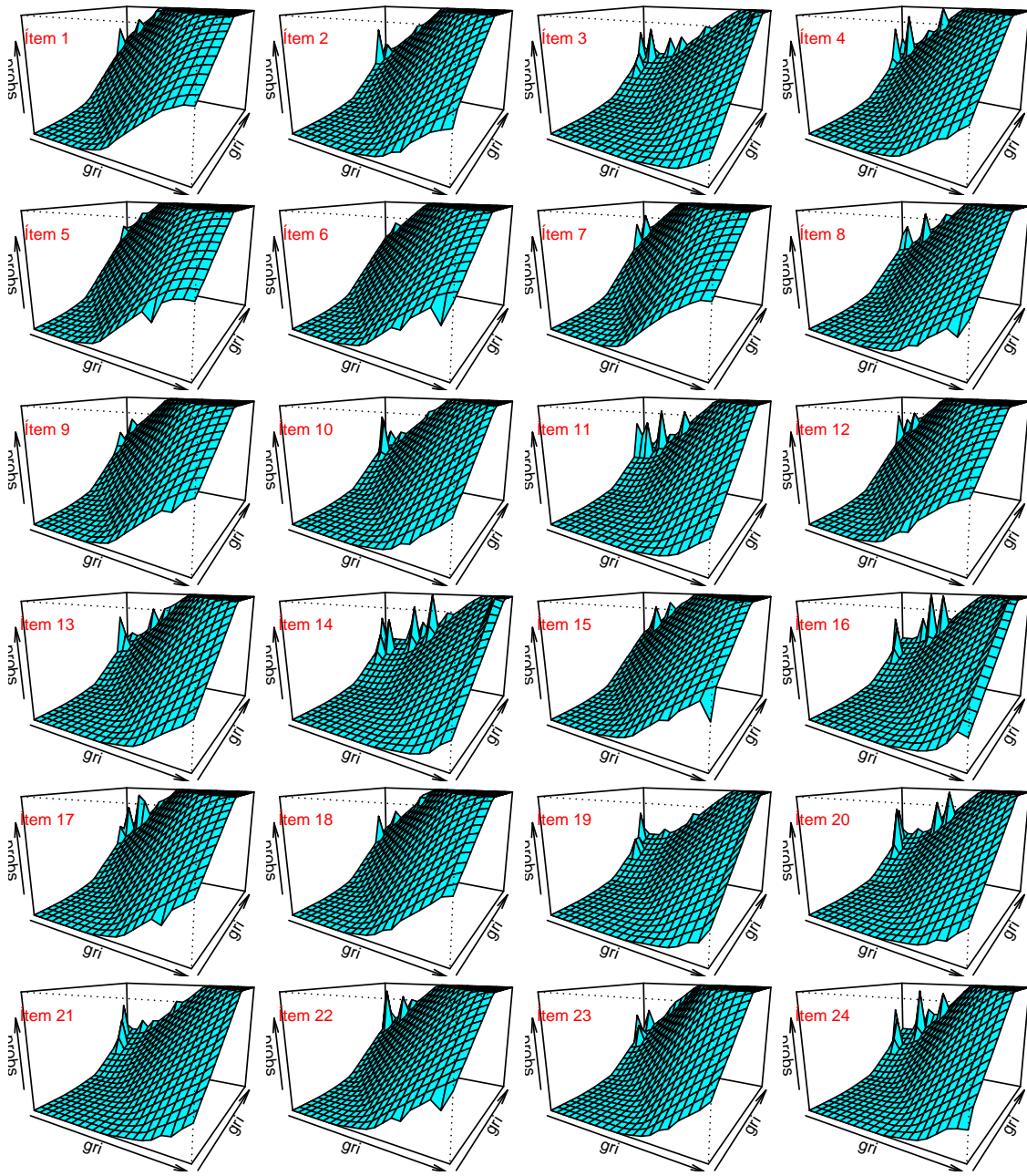


Figura 8.8: Modelo isótono usando regresión no paramétrica polinómica

8.3. FRI de las dimensiones del CDI

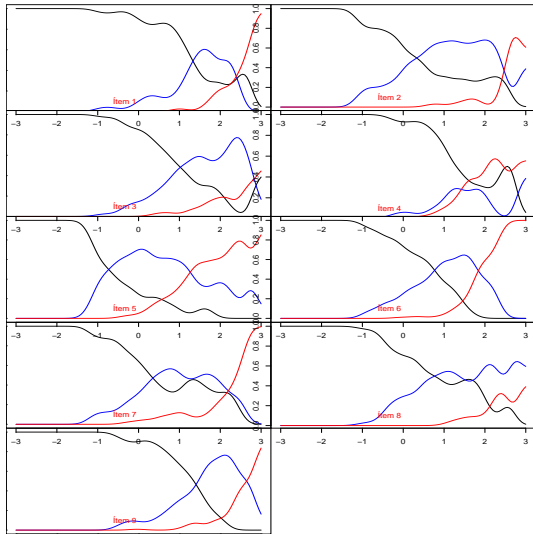


Figura 8.9: Autoestima Negativa Retraimiento
Modelo no paramétrico

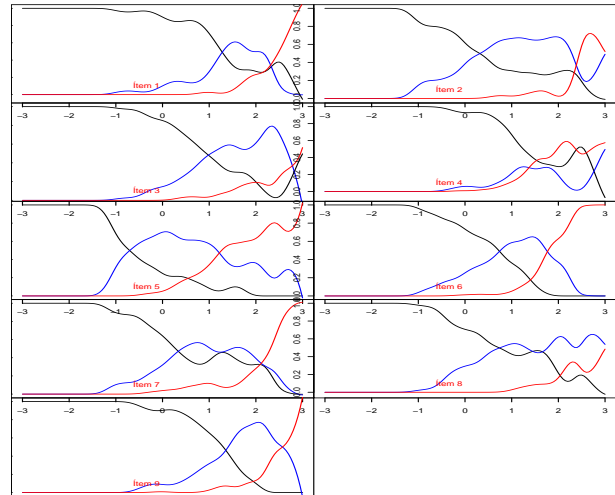


Figura 8.10: Autoestima Negativa Retraimiento
Regresión polinómica

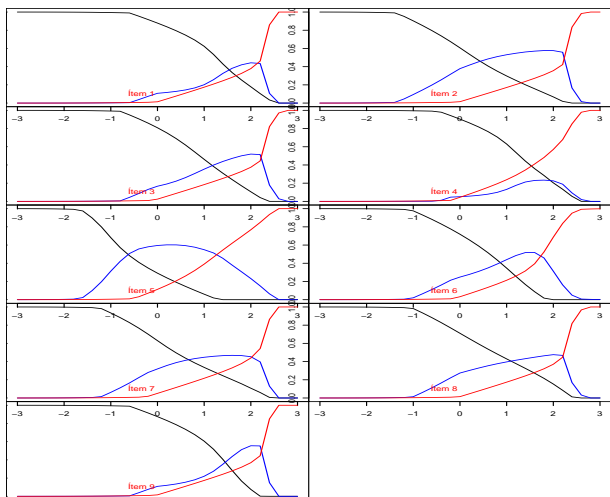


Figura 8.11: Autoestima Negativa Retraimiento
modelo isótono

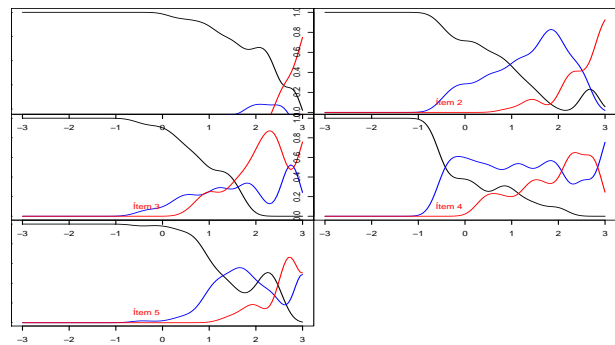


Figura 8.12: Autoestima Negativa Oposiciona-
miento regresión no paramétrica

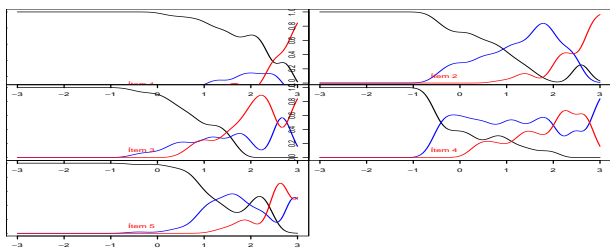


Figura 8.13: Autoestima Negativa Oposiciona-
miento regresión polinómica

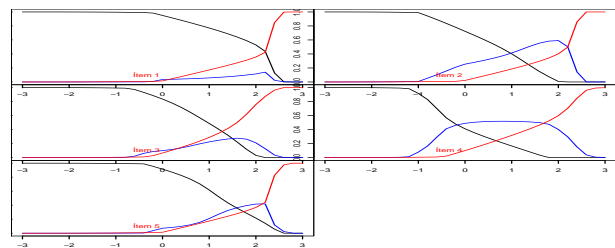


Figura 8.14: Autoestima Negativa Oposiciona-
miento modelo isótono

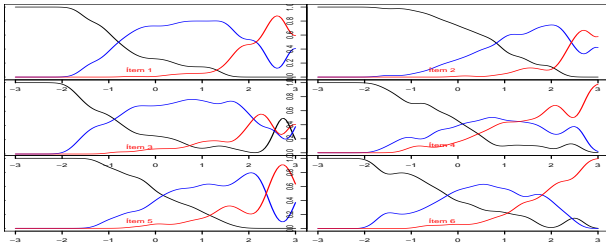


Figura 8.15: Ineficacia regresión no paramétrica

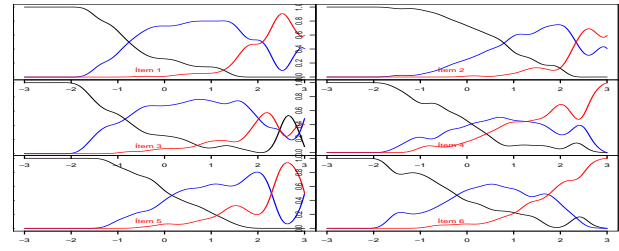


Figura 8.16: Ineficacia regresión polinómica

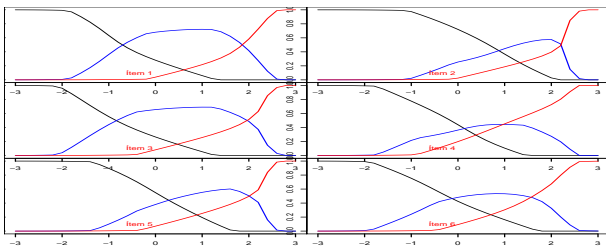


Figura 8.17: Ineficacia modelo isótono

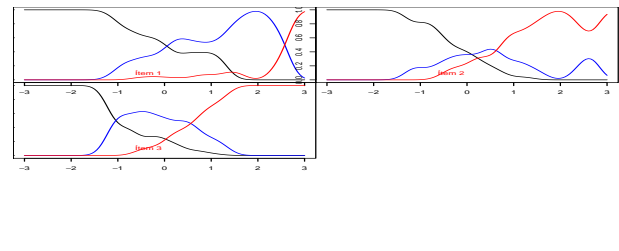


Figura 8.18: Ansiedad regresión no paramétrica

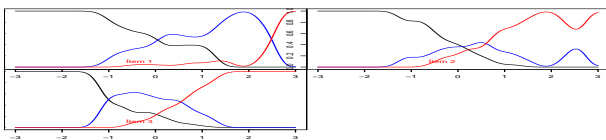


Figura 8.19: Ansiedad regresión polinómica

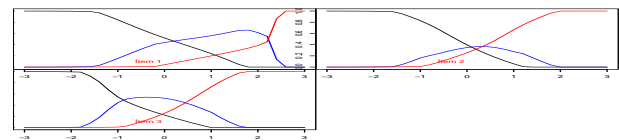


Figura 8.20: Ansiedad modelo isótono

8.4. Proyecciones bidimensionales

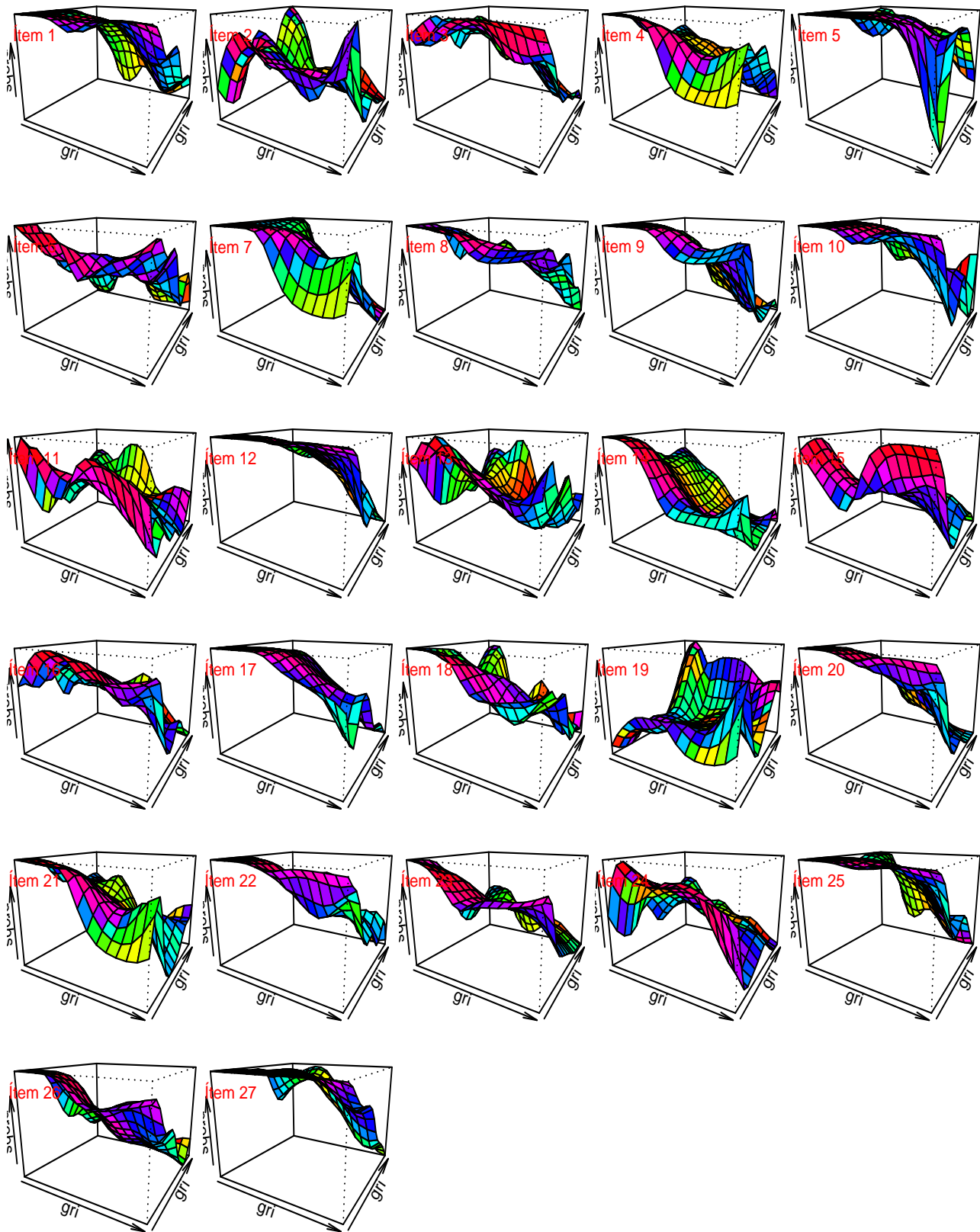


Figura 8.21: Plano Retraimiento-Oposicionamiento (Opción 0)

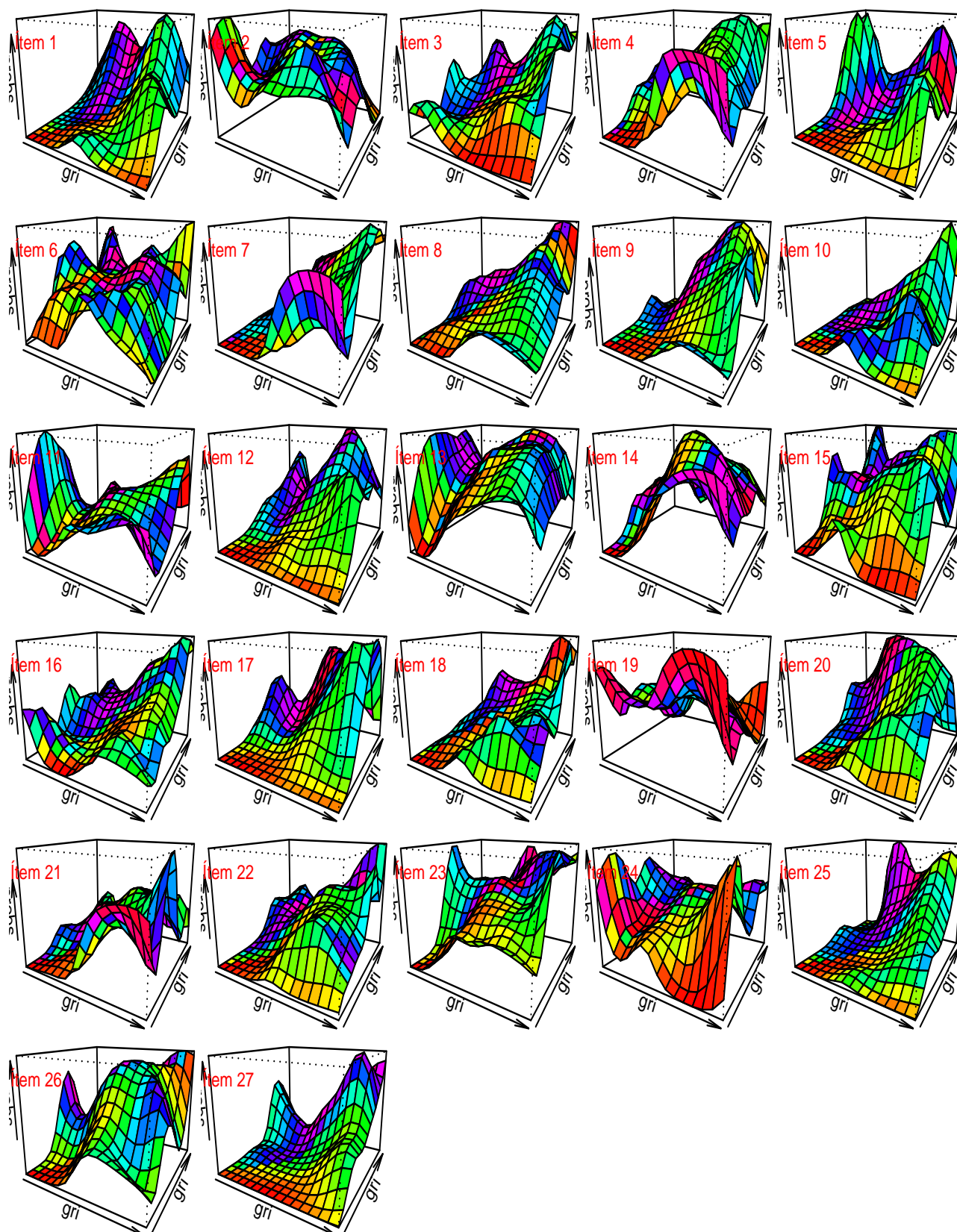


Figura 8.22: Plano Retraimiento-Oposicionamiento (Opción 1)

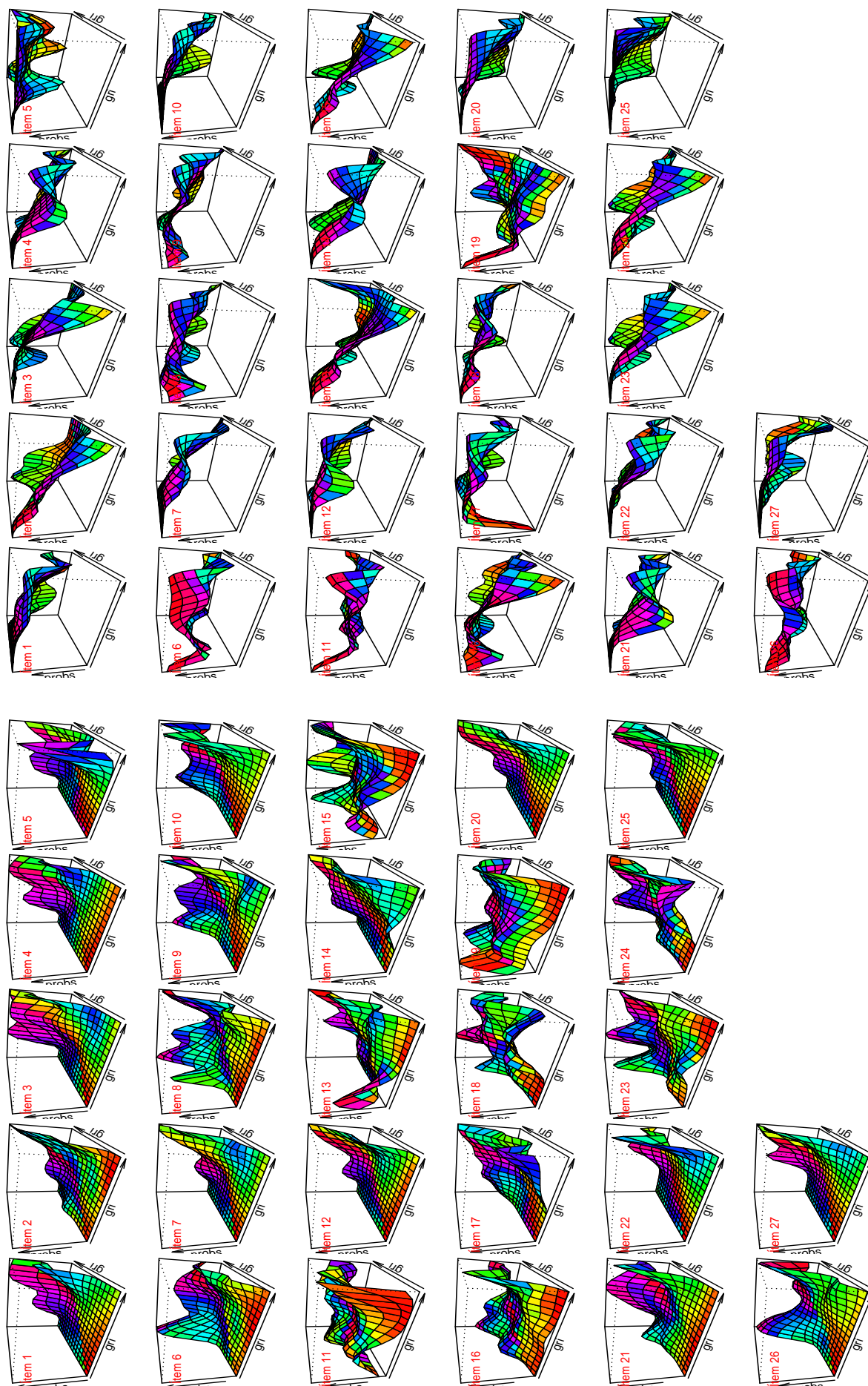


Figura 8.23: Plano Retreinamiento-Oposicionamiento (Opción 2)

Figura 8.24: Plano Retreinamiento-Ineficacia (Opción 0)

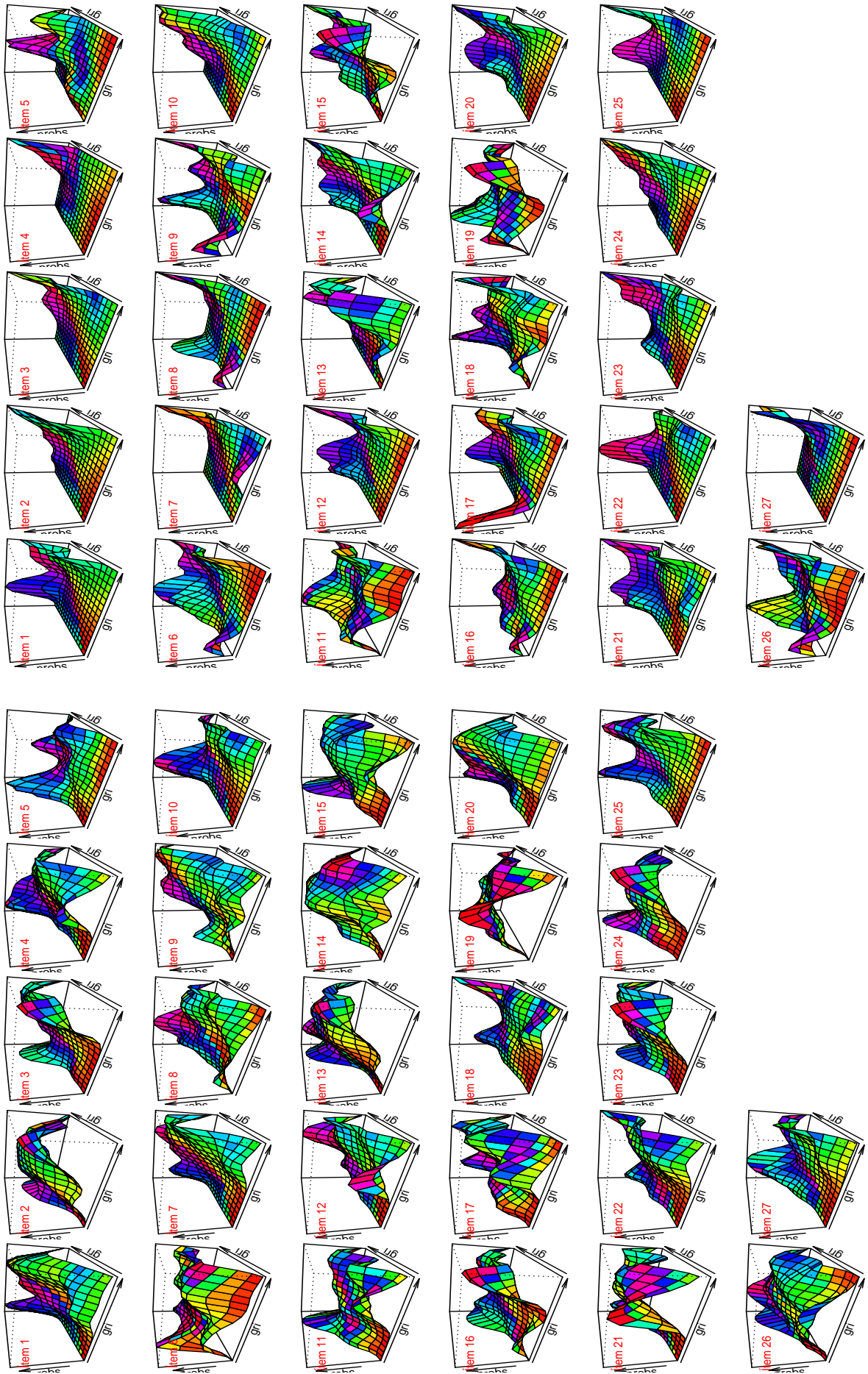


Figura 8.25: Plano Retreinamiento-Ineficacia (Opción 1)

Figura 8.26: Plano Retreinamiento-Ineficacia (Opción 2)

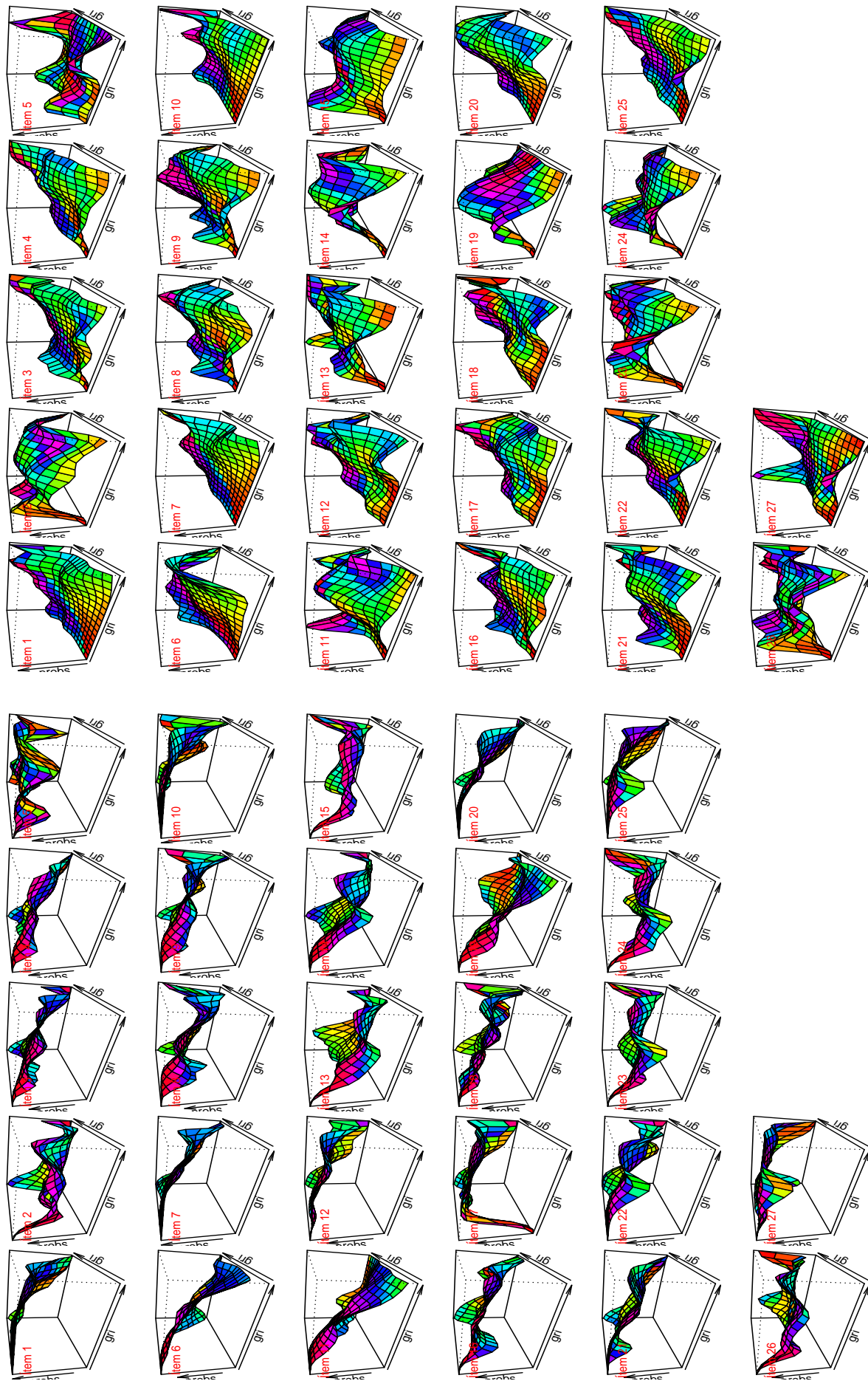


Figura 8.27: Plano Retraimiento-Ansiedad (Opción 0)

Figura 8.28: Plano Retraimiento-Ansiedad (Opción 1)

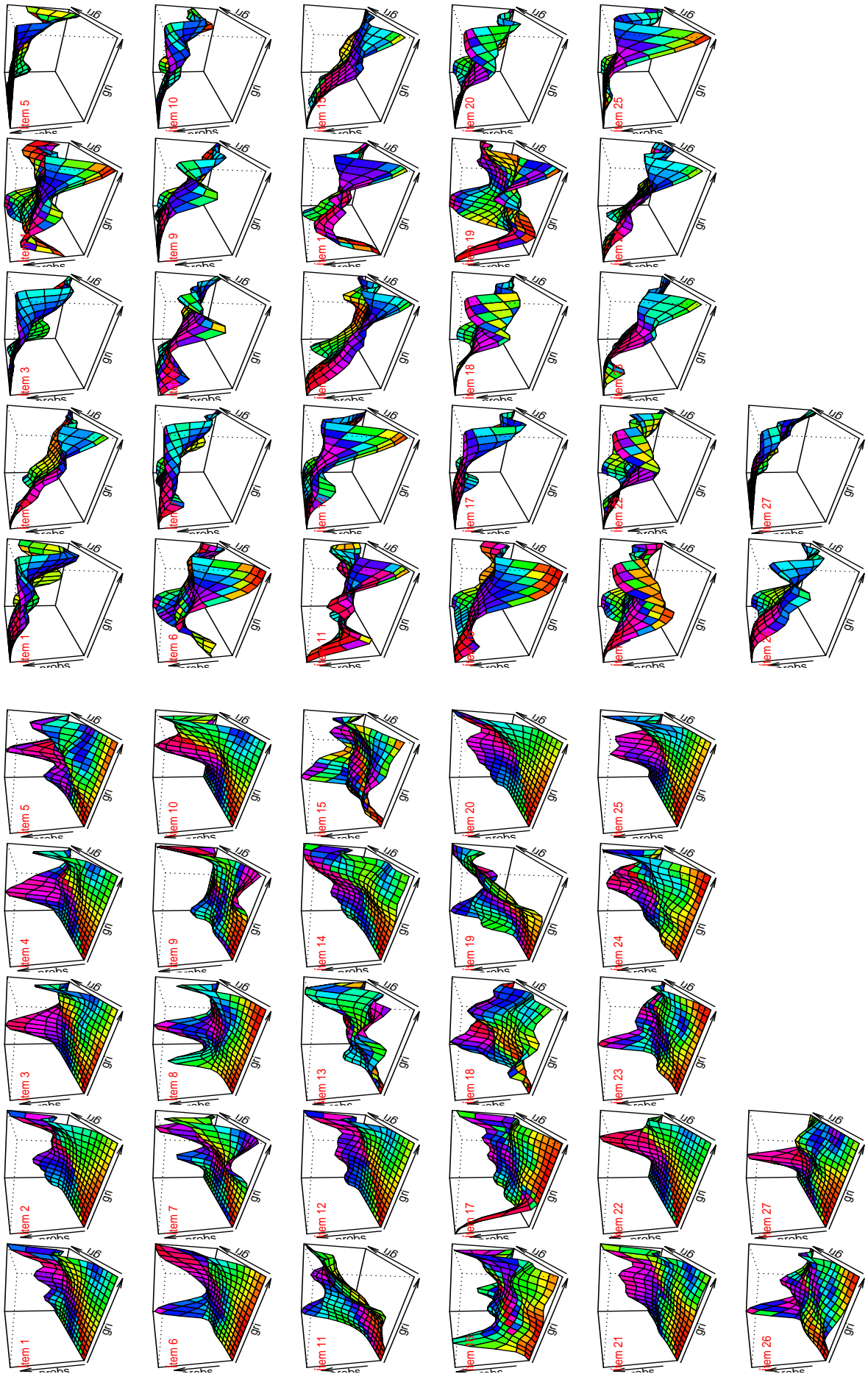


Figura 8.29: Plano Retraimiento-Ansiedad (Opción 2)

Figura 8.30: Plano Oposicionamiento-Ineficacia (Opción 0)

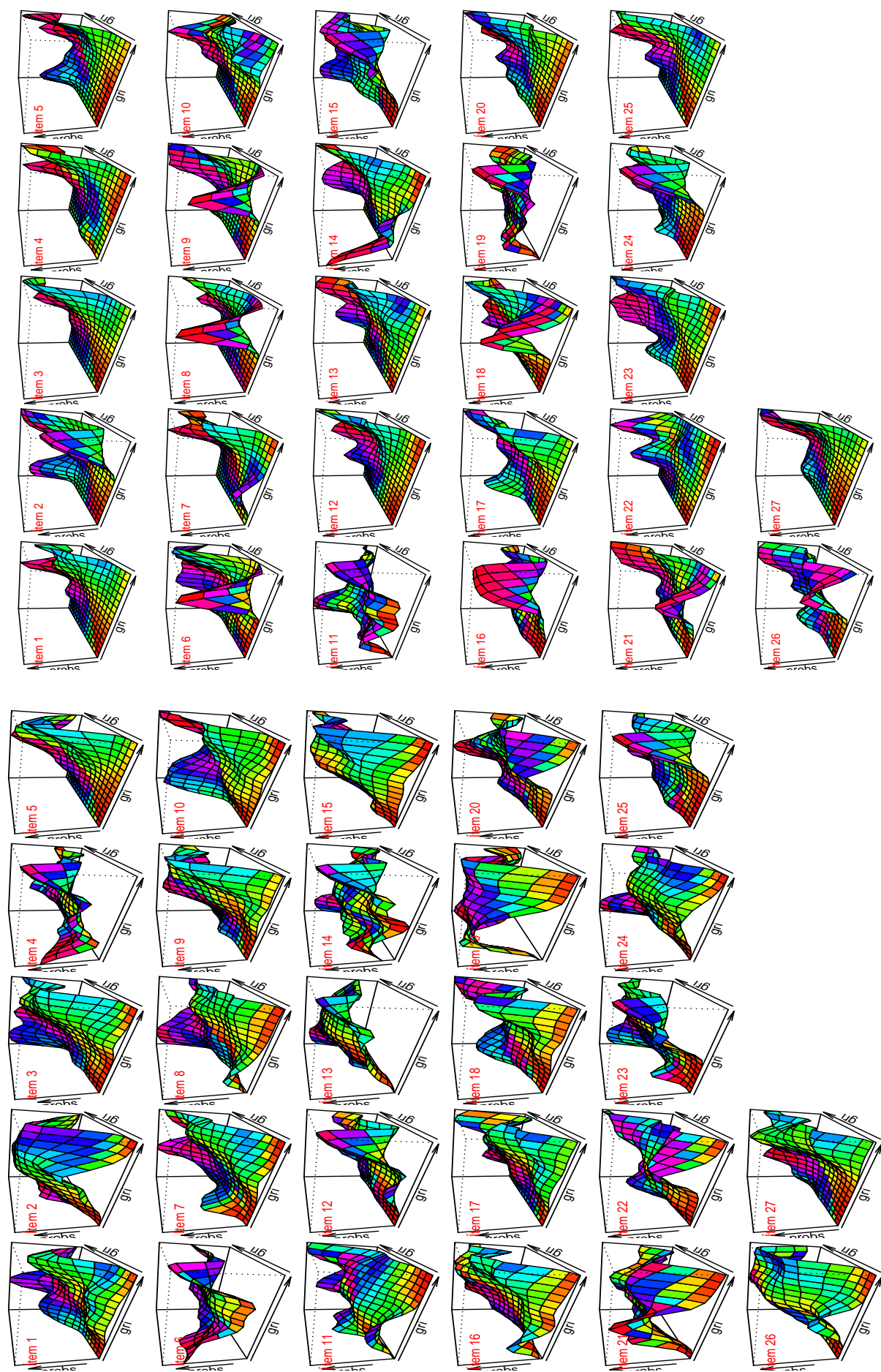


Figura 8.31: Plano Oposicionamiento-Ineficacia (Opción 1)

Figura 8.32: Plano Oposicionamiento-Ineficacia (Opción 2)

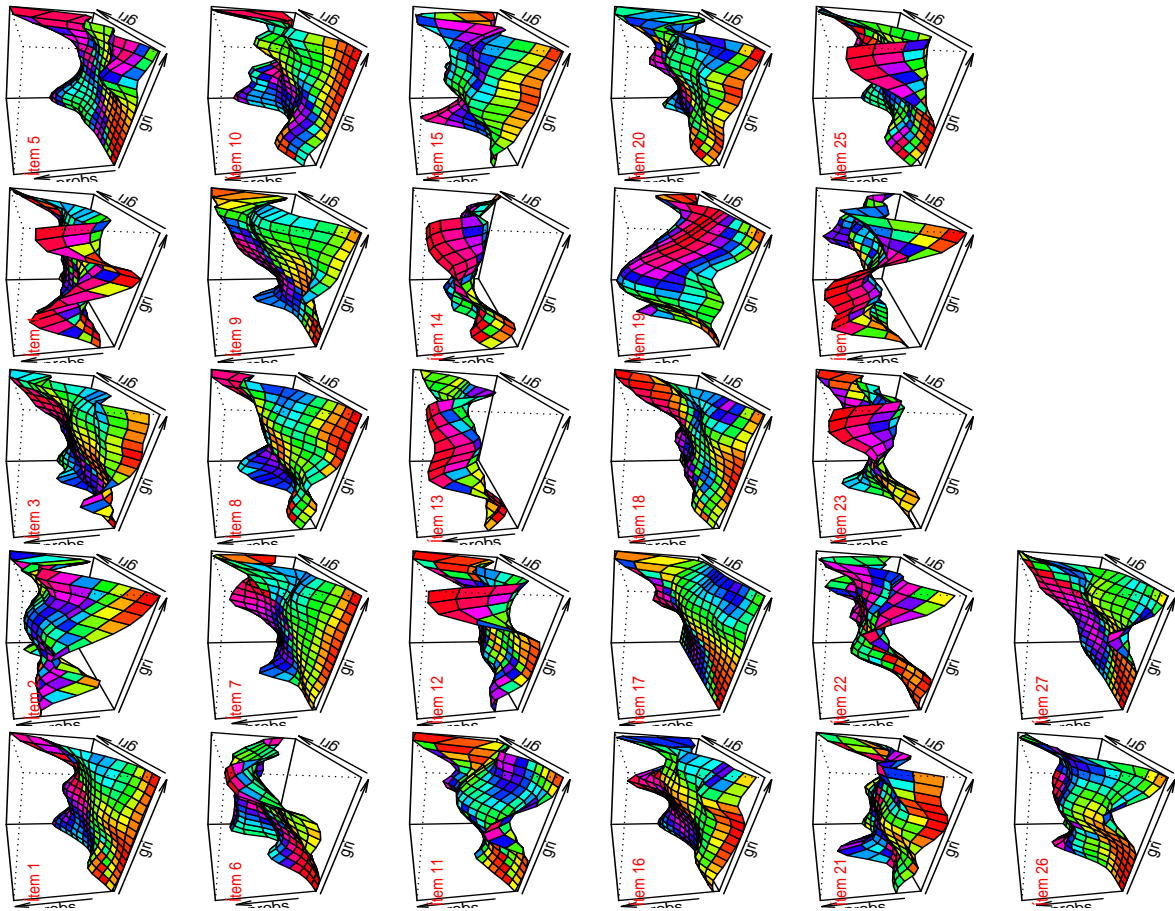


Figura 8.34: Plano Oposicionamiento-Ansiedad (Opción 1)

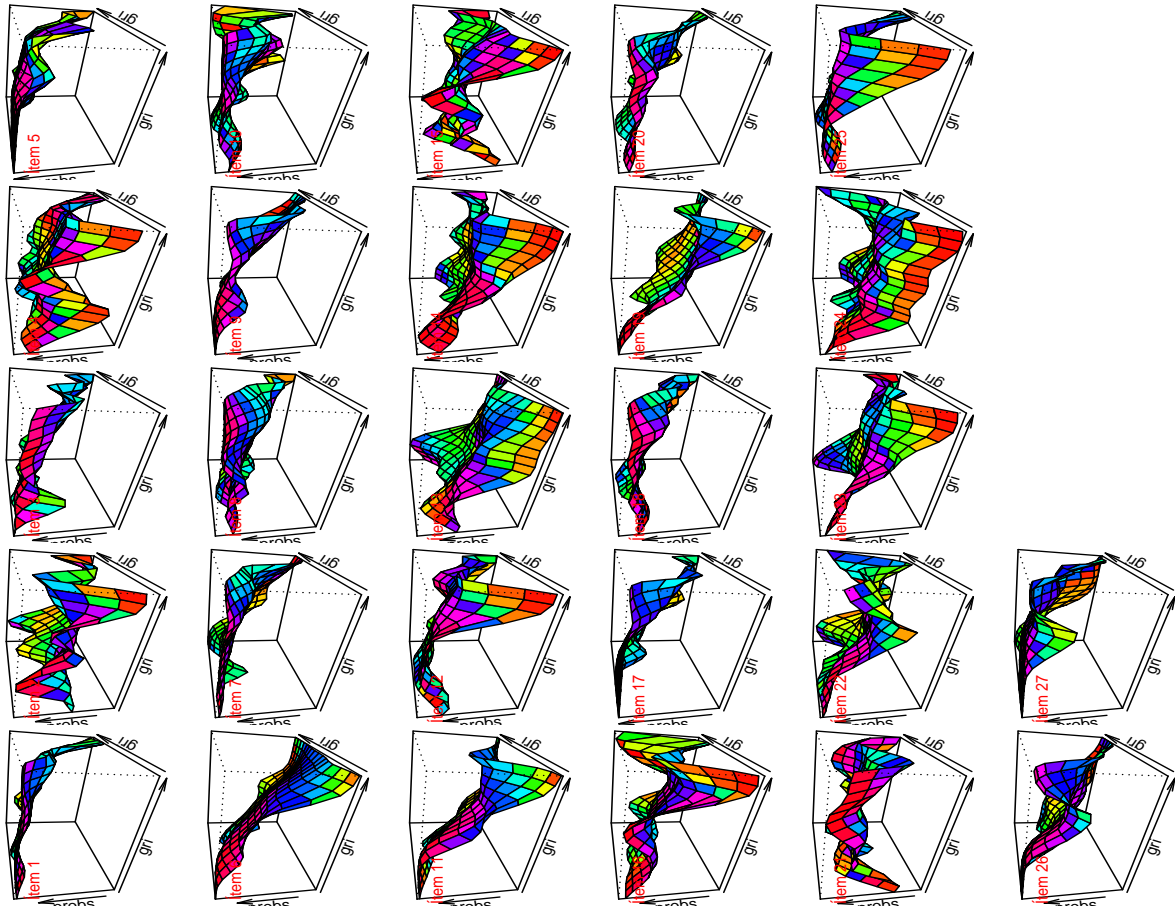


Figura 8.33: Plano Oposicionamiento-Ansiedad (Opción 0)

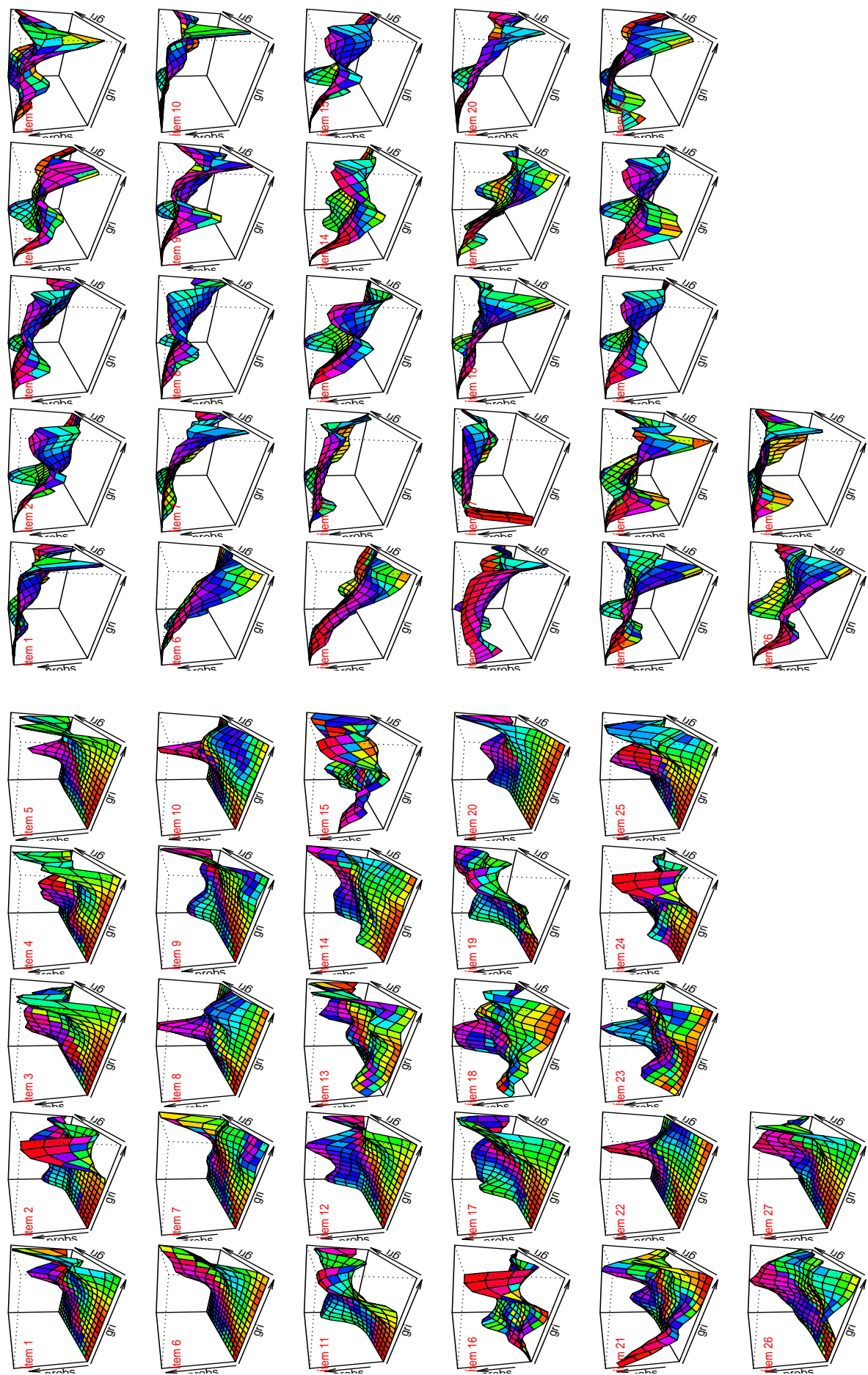


Figura 8.35: Plano Oposicionamiento-Ansiedad (Opción 2)

Figura 8.36: Plano Ineficacia-Ansiedad (Opción 0)

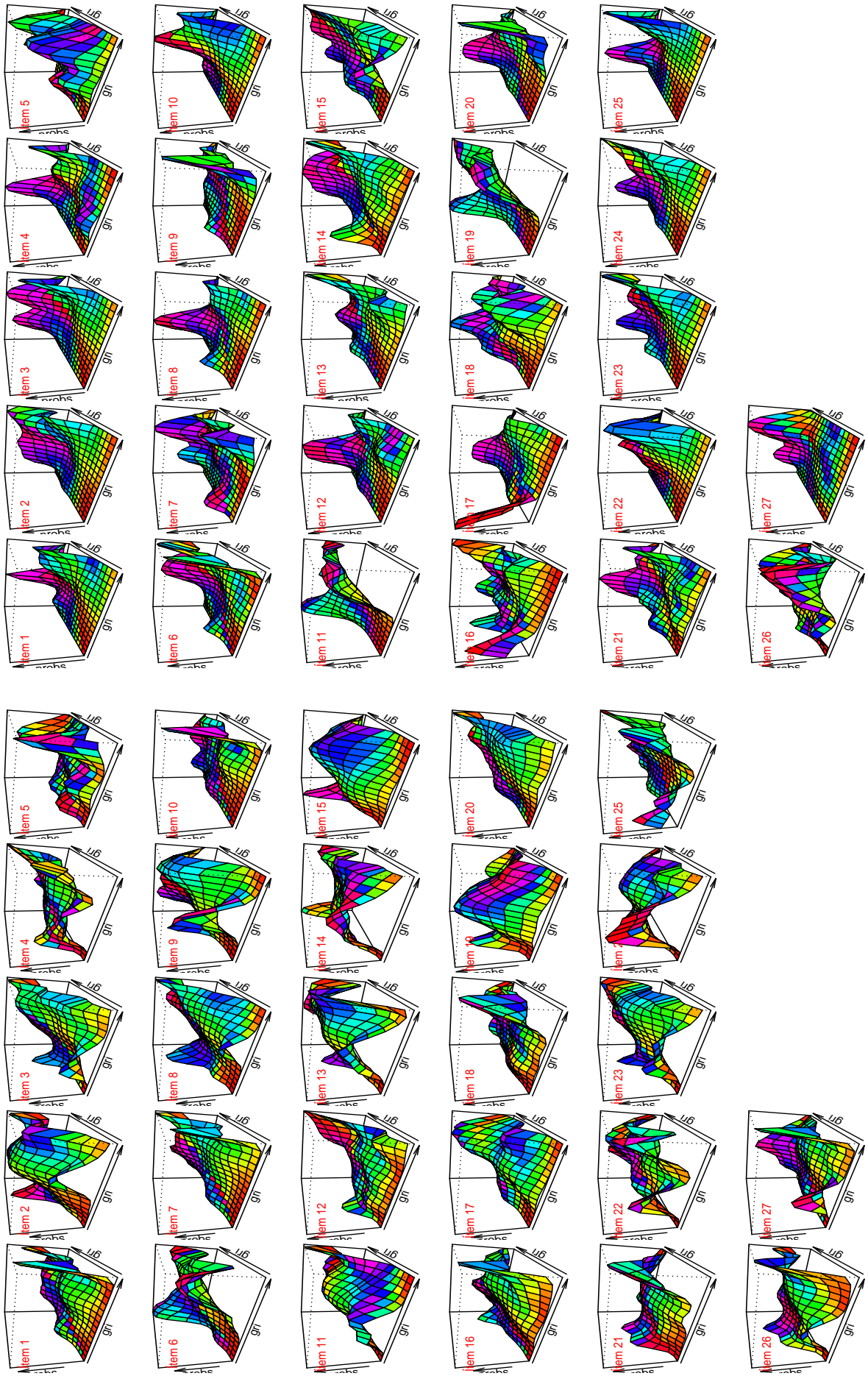


Figura 8.37: Plano Ineficacia-Ansiedad (Opción 1)

Figura 8.38: Plano Ineficacia-Ansiedad (Opción 2)